Autosomal Recessive Retinitis Pigmentosa, Identification and Partial Characterisation of a Novel Gene Implicated in RP25.

A THESIS SUBMITTED TO UNIVERSITY COLLEGE LONDON

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Ciara Áine O'Driscoll
B Sc

DEPARTMENT OF MOLECULAR GENETICS

INSTITUTE OF OPHTHALMOLOGY

UNIVERSITY COLLEGE LONDON

BATH STREET

LONDON

2009

Declaration

I, Ciara Aine O'Driscoll, declare that, unless otherwise stated, this thesis entitled Autosomal Recessive Retinitis Pigmentosa, Identification and Partial Characterisation of a Novel Gene Implicated in RP25 and submitted to University College London for the degree of Doctor of Philosophy is entirely my own original work and composition and has not been submitted for a degree at any other university.

Signed

Ciara Áine O'Driscoll

Date:

ABSTRACT

The purpose of this project is to identify the causative gene for one type of autosomal recessive retinitis pigmentosa, RP25. Through CGH (comparative genome hybridisation) and mutation screening, independent mutations were identified in arRP affected Spanish families mapping to RP25. These mutations were identified within a cluster of uncharacterised gene transcripts all which have EGF-like repeat domains; *Q5T669*, *Q5T1H1, Q9H557_human, Q5TEL3_human, Q5TEL4_human, Q5VVG4_human,* and *Q5T3C8*. Through 5` and 3` RACE PCR analysis, the full length gene was revealed to incorporate the *EGFL11* gene. On assembling all available data we noted that *RP25* gene encompasses 30 exons belonging to nine previously predicted genes and 13 newly identified exons, totaling 43 exons and spanning the interval between 64,487,835 and 66,473,839 on chromosome 6q12.

The *RP25* full length gene transcript is retinal specific. The genomic length covers over 2.0 MB in size and is therefore the largest eye specific gene identified to date. It is also the fifth largest gene in the human genome to date. Homologs of the *RP25* gene to Drosophila *eys/eys-shut* (Spacemaker) were identified, leading to the annotation of the name *EYS* (SPAM). An apparently intact eys gene is found across the mammalian clade, including monotremes (platypus) and marsupials (opossum). However, despite the mutations and the presumed loss of function associated with human disease, this gene has been dispensed with on at least four separate occasions in the last 100 million years of mammalian evolution including in the armadillo (*Dasypus novemcinctus*), little brown bat (*Myotis lucifugus*) and ruminant (cattle and sheep) lineages. *EYS* has acquired several (<3) reading-frame disruptions in three rodents (mouse, rat and guinea pig) representing

two of the three major rodent clades. Through immunohistochemical and electron microscopy analysis, a signal for SPAM was identified in the outer segments of photoreceptor cells.

Acknowledgments

I would like to thank Professor Shomi Bhattacharya for being a wonderful supervisor and mentor for the past few years. He has also witnessed his fair share of my tears and has never once discouraged me or my work. He has boosted my confidence in my ability and without his dedicated support I would never have been able to complete this PhD thesis. I would like to thank Mai, who has been an immense support, inspiration and guide through my PhD years. It is an understatement to say that without Mai this thesis would not have been completed to the standard it is now. I would also like to thank our collaborators in Spain, Guillermo and Isabella who have both been wonderful to work with as well as fantastic friends.

I would also like to thank the girls in the office, especially Francesca, Amna, Christina, Sancy, Brotati, and Kinga who have all been there through the ups and downs. These are the girls who have made my time special through the last three years. I must also acknowledge Wayne who also has been an amazing support and friend to me during my studies. I would also like to thank Quincy, Rebecca, Gio, Vinny, Bev and Naushin all of whom have offered great advice and support. I would like to thank the porters who welcomed me with a smile and some banter every day. I must also thank Peter Munro for his time and patience when teaching me EM. I would also like to thank Peter Marshall who made sure I was funded all of the time. I must also acknowledge all of the friends I made during my Neurotrain participation; I believe I have made some friends for life from this network.

Dedication

I would like to dedicate this thesis to my family, who over the past three years of hair pulling, tear blubbing and tantrum making have been so supportive throughout. My parents have been my inspiration and my sanity. In my times of desperation they have calmed me down and listened to my every gripe and moan. I would like to thank my sisters Sinead, Aoife, Muireann and my brother Damien, who have all been so supportive and proud.

List of Publications

Abd El-Aziz MM*, Barragan I*, **O'Driscoll CA***, Goodstadt L, Prigmore E, Borrego S, Mena M, Pieras JI, El-Ashry MF, Safieh LA, Shah A, Cheetham ME, Carter NP, Chakarova C, Ponting CP, Bhattacharya SS, Antinolo G. (2008) EYS, encoding an ortholog of Drosophila spacemaker, is mutated in autosomal recessive retinitis pigmentosa. *Nat Genet.* **40**(11):1285-7.

Barragán I, Abd El-Aziz MM, Borrego S, El-Ashry MF, **O'Driscoll C**, Bhattacharya SS, Antiñolo G. (2008) Linkage validation of RP25 Using the 10K genechip array and further refinement of the locus by new linked families. *Ann Hum Genet.* **72**(4):454-62.

Abd El-Aziz MM, Barragan I**, O'Driscoll C**, Borrego S, Abu-Safieh L, Pieras JI, El-Ashry MF, Prigmore E, Carter N, Antinolo G, Bhattacharya SS. (2008) Large-scale molecular analysis of a 34 Mb interval on chromosome 6q: major refinement of the RP25 interval. *Ann Hum Genet.* **72**(4):463-77.

Presentations

O'Driscoll C.A. (2008) Development of Coding SNPs as Markers for Retinal Dystrophy Loci on Chromosome 6q. ARVO E-abstract (1303/A295).

Table of Contents

List of figures

List of Tables

LIST OF ABBREVIATIONS:

A: adenine

ABI: Applied Biosystems Incorporated

ad: Autosomal dominant

AMD: Autosomal macular degeneration

ar: autosomal recessive

AR: antigen retrieval

bp: base pair

BLAT: BLAST like alignment tool.

BLAST: Basic local alignment search tool.

C: Cytosine

Ca^{2+}: Calcium

CC: Connecting Cilium

cDNA: complementary deoxyribonucleic acid

CGH: Comparative genome hybridization

cGMP: Cyclic guanine monophosphate

cGMP-PDE: cGMP-phosphodiesterase

CNVs: Copy number variations

cM: CentiMorgan

CORD: Cone-rod dystrophy

DAPI: 4'-6-Diamidino-2-phenylindole

DNA: Deoxyribonucleic acid

dNTP: Deoxynucleotide triphosphate

EDTA: Ethylenediamine-tetra acetic acid

EGF: Epidermal growth factor

FITC- Fluorescein isothiocyanate

G: Guanine

GCL- Ganglion Cell Layer

HIER- heat induced epitope retrieval

HMMER- Hidden Markov Models

INL: Inner nuclear layer

IS: Inner Segment

Kb: Kilobase

LCA: Leber congenital amarousis

LD: Linkage disequilibrium

MAPH: multiplex amplifiable probe hybridization

Mb: Megabase

MDCK: Madin-Darby canine kidney

MLPA: multiplex ligation probe dependent assay

mRNA: messenger ribonucleic acid

NCBI: National Centre of Biotechnology Information

NEIBank: The National Eye Institute Bank

ONL: Outer nuclear layer

ORF: Open reading frame

OS: Outer segment

PBS: Phosphate buffered saline

PCR: Polymerase Chain Reaction

PFA-paraformaldehyde

PIER- proteolytic induced epitope retrieval

PNA: Peanut Agglutinin

PHYLIP: *PHYL*ogeny *I*nference *P*ackage

RACE: Rapid Amplification of cDNA ends

RFLP: Restriction Fragment Length Polymorphism

RHO- rhodopsin

RPE: Retinal pigment epithelium

RNA: Ribonucleic acid

ROS: Rod Outer segment

RIS: Rod Inner Segment

SNPs: Single Nucleotide Polymorphism

T: Thymine

TRITC tetramethylrhodamine

WGA: Wheatgermagglutinin

WGTP : whole genome tiling path

CHAPTER 1

INTRODUCTION

1.1 The human eye

The human eye is the organ which reacts to light and creates sight. The eyeball is set in a protective cavity in the skull called the orbit or socket. The eye is made up of three layers which enclose three transparent structures.

The outer layer is made of two parts: the anterior cornea and the posterior sclera. The cornea is transparent and allows entry of light into the eye; the sclera creates the "white" of the eye and is composed of fibrous tissue (Atchison and Smith, 2000). The middle layer is divided into three structures: the iris, the ciliary body and the choroid. The iris is the coloured part of the eye, and it controls light levels. The ciliary body produces aqueous humour, this clear fluid fills the front of the eye. The final segment of the middle layer is the choroid; this is the main source of blood in the eye. The innermost layer of the eye is the retina, which is an extension of the central nervous system and is connected to the brain through the optic nerve (Atchison and Smith, 2000). The inside of the eye is divided into three segments. This includes the vitreous cavity which is between the lens and the retina and contains a transparent mass called the vitreous humour (Figure 1.1). The layer of most interest in this study is the retina.

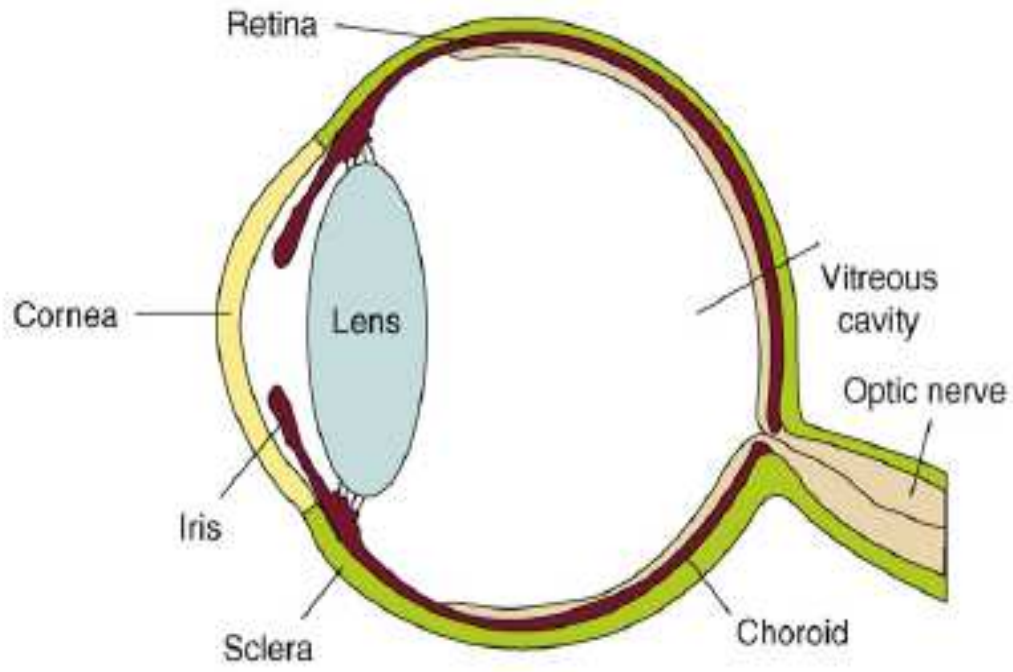


Figure 1.1: A schematic view of the eye. The structure of the eye is made up of various layers of tissue, starting from the outset the conjunctiva, all the way through to the retina (Smith *et al.,* 2009).

1.2 The Retina

The retina is a specialised sensory cell layer that is capable of transforming light into an electrical signal. This signal is transmitted via the optic nerve to the visual centers of the brain.

The retina is composed of 10 layers, from the outside (nearest the choroid) to the inside (nearest the vitreous humour) (Figure 1.2)

Pigmented epithelium

Photoreceptors

External (outer) limiting membrane

Outer nuclear layer

Outer plexiform layer

Inner nuclear layer

Inner plexiform layer

Ganglion cells

Nerve Fiber layer

Internal limiting membrane

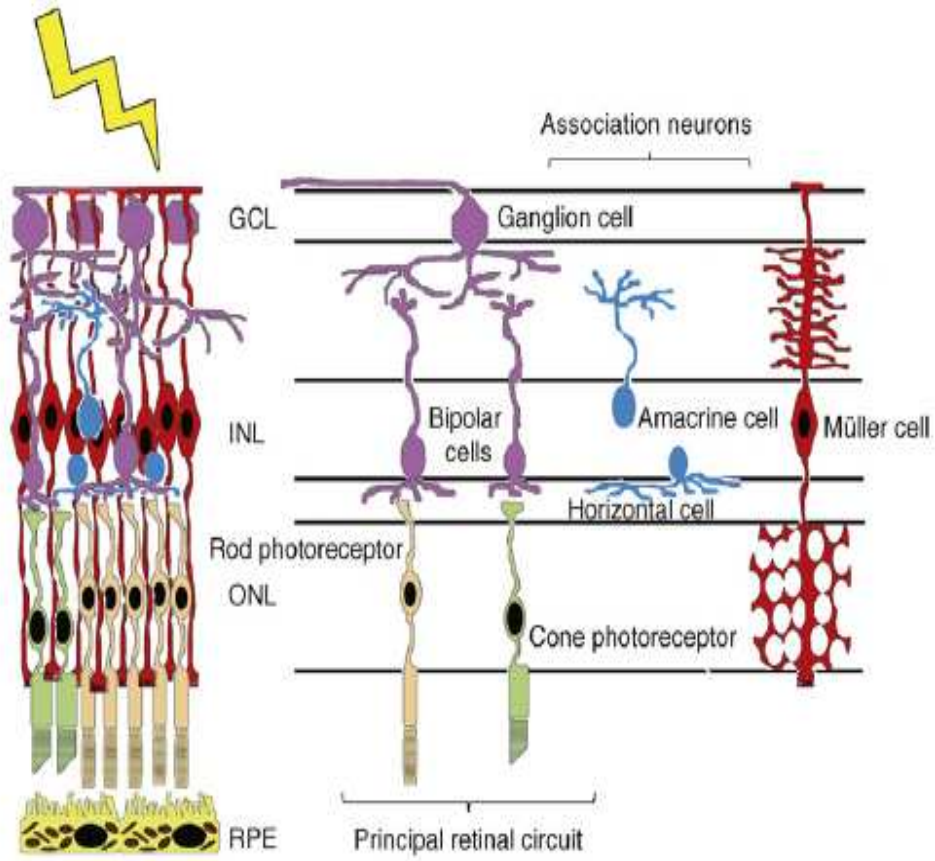


Figure 1.2: Layers and cell types of the retina. The direction of light through the cell layers is depicted by the arrow (Smith *et al.,* 2009).

1.2.1 Pigmented Epithelium

The retinal pigment epithelium (RPE) is a continuous monolayer of cuboidal epithelial cells that is densely packed with pigment granules (Schraermeyer and Heimann, 1999). This cell layer has many physical, optical, metabolic/biochemical and transport functions. These roles include maintaining adhesion of the retina, providing a selectively permeable membrane between the choroid and the retina and absorbing light. It nourishes retinal visual cells, and is firmly attached to the underlying retinal visual cells. The RPE consists of a single layer of approximately 4-6 million cells firmly attached to its basal lamina. The RPE is involved in the phagocytosis of the outer segment of photoreceptor cells. It is also involved in the vitamin A cycle; it isomerizes trans retinol to 11-cis retinal (Katz and Gao, 1995). It also maintains the retinal environment by supplying small molecules such as D-glucose. Interdigitations between the RPE microvillus and the photoreceptor outer segment provide a physical closeness between the two layers and the material that occupies the space between the two layers called the interphotoreceptor matrix (IPM) provides adhesive forces (Hageman *et al.,* 1995).

1.2.2 Photoreceptors

Photoreceptors are the cells responsible for converting light into nerve signals that will be transmitted to the brain via the optic nerve. They send signals to other neurons by a change in their membrane potential when they absorb photons. This process is known as signal transduction. In vertebrates, there are two types of photoreceptor cells: rods and cones. The human retina has 120 million rods and 6 million cones (Koenekoop 2009). Photopigment or coloured material fill the rods and cones and absorb particles

of light that strike the retina. A photoreceptor cell contains a membranous photoreceptor protein called an opsin which contains a pigment molecule called retinal.

Both types of photoreceptors have the same basic structure. Closest to the visual field (and furthest from the brain) is the axon terminal, which releases a neurotransmitter called glutamate to bipolar cells. Further back is the cell body, which contains the cell's organelles. Further back still is the inner segment, a specialized part of the cell full of mitochondria. The chief function of the inner segment is to provide ATP (energy) for the sodium-potassium pump. Finally, closest to the brain (and furthest from the visual field) is the outer segment, the part of the photoreceptor that actually absorbs light. Outer segments are actually modified cilia that contain discs filled with opsin, the molecule that actually absorbs photons, as well as voltage-gated sodium channels (figure 1.3). Detailed structure of Rods and Cones will be described under the following subsection:

1.2.2.1 Rods

Rods are long (100-120μm) slender cells whose outer segment contains the visual pigment rhodopsin within the membrane-bound lamellae that are enclosed by a single cell membrane (Luo *et al.,* 2008). The rod-shaped outer segment of a rod cell, which is the light-sensitive part, consists largely of a stack of flat and parallel unit membranes called discs. Discs are enclosed by the apical microvilli of the RPE. Rods are very sensitive to light nerve cells, because of which these cells Discs are enclosed by the apical microvilli of the RPE. Rods are very sensitive to light nerve cells, because of

which these cells are responsible for night vision. The pigment in the rods is called rhodopsin or visual purple, and enables the eye to see shades of gray and see in dim light (figure 1.4).

1.2.2.2 Cones

Cones are adapted to detect colors, and function well in bright light. They have a similar structure to rod cells but their outer segment which contains the photoreceptive pigments is cone shaped. Cones require a reasonable amount of light to be stimulated so only function during the day. This is why we have colour vision but mainly during the daytime when the light intensity is high enough to stimulate the cone receptors. Like rods, cones contain an opsin and the chromophore 11-cis-retinal (Reme and Wenzel, 2003). There are three types of pigment or opsins in cones: cyanolabe absorbs blue light, chlorabe absorbs green light and erythrolabe absorbs red light. The wavelengths of these opsins are at 560nm, 426nm and 530nm respectively which allow the eye to visualize colors and to detect sharp images in bright light. Cones are about 60-75 μm in length and are spread throughout the retina. The discs in cones are not surrounded by a plasma membrane in the same manner as rods. Cones have a longer lifespan than rods and do not undergo circadian phagocytosis by the RPE cells (Schraermeyer and Heimann, 1999).

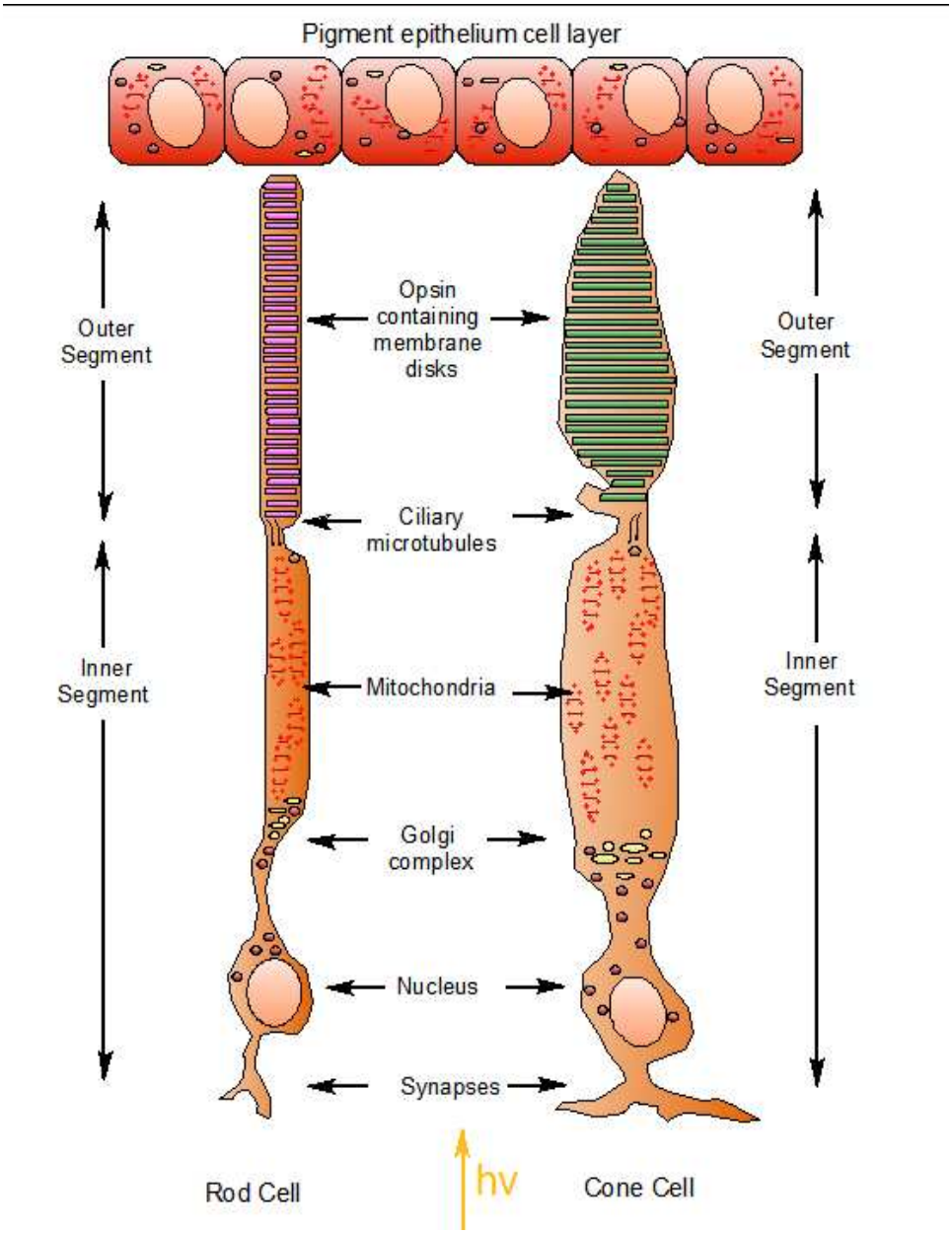


Figure 1.3: Structure of rod and cone photoreceptors, including the outer and inner segments. Direction of light hitting photoreceptors is depicted with the yellow arrow. http://www.ganfyd.org.

1.2.3 Outer limiting membrane, outer nuclear later, outer plexiform layer

The outer limiting membrane is the site of cut off junctions which seal off the disposable parts of rods and cones. This layer has the false appearance of a "membrane" in light microscopic preparations, due to the presence of many junctions tightly packed together. The outer limiting membrane acts as a barrier to keep the inner layers of the retina safe from harmful material.

The outer nuclear layer contains the nuclei and cell bodies of rods and cones. The outer plexiform layer is a layer of neuronal synapses. It is composed of a network of synapses between dendrites of horizontal cells from the inner nuclear layer and photoreceptor cell inner segments from the outer nuclear layer. Horizontal cells span across cones and take inputs from them to control the amount of GABA (amino butyric acid) released back onto photoreceptor cells. There are three basic types of horizontal cells, designated HI, HII and HIII (Kolb *et al.,* 1992).

1.2.4 Inner nuclear layer, Inner plexiform layer, Ganglion cells, Nerve Fiber layer, Internal limiting membrane

The cell bodies and nuclei of the integrator neurons, particularly the bipolar and horizontal cells, are located in the inner nuclear layer, next inwards from the outer plexiform layer. The retina contains approximately 35.7 million bipolar cells which comprise of several functional and morphologic subtypes. Bipolar cells relay information from photoreceptor cells to horizontal and ganglion cells and receive extensive synaptic feedback from amacrine cells (Ayoub and Matthews, 1992). Their

cell bodies lie in the inner nuclear layer and are oriented in a radial fashion parallel to the photoreceptors. Bipolar cells can be divided into nine subtypes: one rod bipolar type and 8 cone bipolar cell types.

The inner plexiform layer (IPI), like the outer one, is a region of synapses. Here the bipolar cell processes synapse with the dendritic processes of ganglion cells, the final neuronal element of the eye. These are interneurons which deliver ganglion cell input and regulate output of cone bipolar cells. They operate in the IPL of which there are approximately 40 subtypes, most lacking axons (Wassle and Boycott, 1991). Like other neurons, these ganglion cells have axons, which carry the generated signal. The axonal fibers from the ganglion cells, bundled together, run radially around the inner surface of the retina, converging at the site of origin of the optic nerve. At that point the nerve penetrates the retina and the sclera. Here there are no light sensitive elements; this is known as “the blind spot” in vertebrates.

The innermost layer is not really a membrane, but is instead a place of fused cell processes. The fusion is between foot processes of the Muller cells. This fusion completes the function of sealing off the retinal components from the potentially harmful materials present in the vitreous chamber. Like the outer limiting membrane, it is a region of occluding junctions and a "tight" barrier which can be compared to that formed by glial cells in the brain.

1.3 Phototransduction Cascade

Phototransduction is the process of converting light energy into electrical signals via photoreceptor cells. The components of this cascade are uniquely expressed in photoreceptors and, although functionally similar, distinct components of these cascades are expressed in rods and cones. The molecular reactions of phototransduction can be separated into two major divisions (i) activation and (ii) termination and modulation (Lamb and Pugh, 2006). Activation refers to the steps that lead to the onset of the light response. Termination and modulation refer to the steps that lead to the shut-off of the activated molecule. Light initiates the isomerization of 11-cis retinal to all-trans retinal resulting in the activated state of rhodopsin (RHO), also known as RHO*. RHO* is vital for the phototransduction cascade where it initiates the visual cycle. RHO* triggers a G-protein (G) coupled enzymatic cascade catalyzing the exchange of GDP for GTP, and converts transducin into an active form (T*). Activated transducin in turn activates cGMP-phosphodiesterase (cGMP-PDE), which brings about the hydrolysis of channel bound cGMP to 5'-GMP. The resulting rapid decrease in intracellular cGMP levels causes the cGMP-gated channels to close and the rod cell becomes hyperpolarized. Termination of the light-response cascade is mediated by inactivation of RHO* by phosphorylation catalysed by rhodopsin kinase and is then bound by arrestin, which prevents further interaction with transducin. Transducin and phosphodesetrase (PDE) are inactivated by the hydrolysis of GTP to GDP. Guanylate cyclase, the enzyme responsible for the synthesis of cGMP from GTP, is inactivated by the decrease in intracellular Ca^{2+} after photoexcitation. As the cGMP concentration

increases, the cGMP-gated channels re-open, and the photoreceptor is returned to its depolarized state (Figure 1.4).

1.4 The Visual Cycle

The visual cycle which is also known as the retinoid cycle is the regeneration of the visual pigment. Light perception in vertebrates is mediated by a group of G protein–coupled receptors called opsins (e.g. rhodopsin), which are bound to a chromophore (light-absorbing compound), 11-*cis*-retinal, derived from vitamin A (Johnson *et al.,* 2003). Absorption of light induces an 11-*cis*-retinal to all-*trans*-retinal isomerisation, and the conformational change initiates the visual transduction cascade. In order for the activated opsin to participate again in the visual process, the all-*trans*-retinal must be isomerised back to the 11-*cis* form via an enzymatic pathway, the visual cycle (Figure 1.5) which occurs mainly within the RPE. Following isomerization and release from the opsin protein, all-*trans* retinal is reduced to all-*trans* retinol and travels back to the RPE to be "recharged". It is first esterified by lecithin-retinol acyltransferase (LRAT) and then converted to 11-*cis* retinol by the isomerohydrolase RPE65. Finally, it is oxidized to 11-*cis*-retinal before traveling back to the rod outer segment where it can again be conjugated to an opsin to form a new, functional visual pigment (rhodopsin). Any protein or enzyme mutated at any stage of the visual cycle or the phototransduction cascade can lead to different types of retinal degenerations. These will be discussed in detail in the following sections.

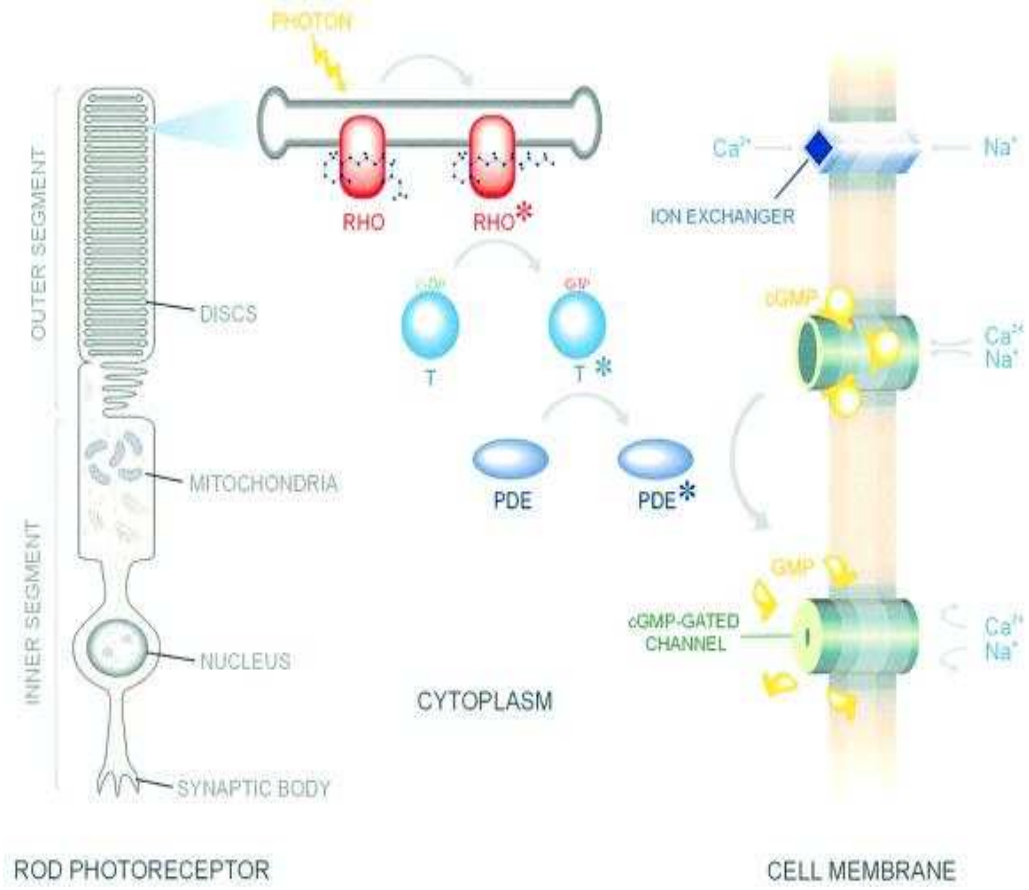


Figure 1.4 Phototransduction cascade- Overview of the events of the phototransduction pathway, from the initiation of the photon of light all the way through to the depolarization of the photoreceptor (Farrar *et al.,* 2002).

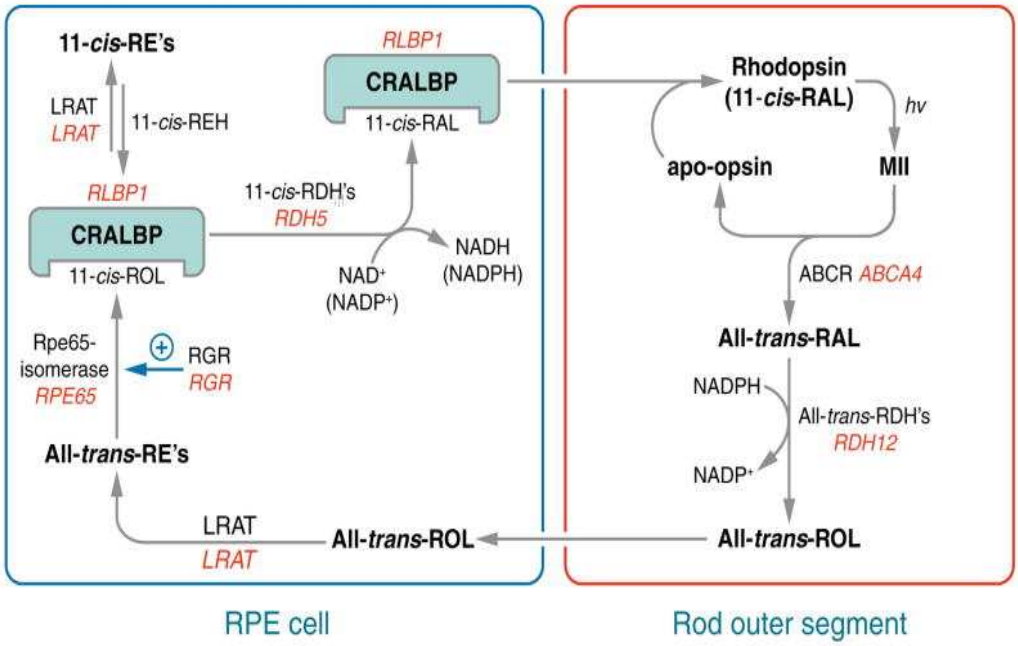


Figure 1.5 The visual cycle. Light induced rhodopsin (RhO) produces a photoactivated metrhodopsin II (R*), which converts to opsin and all-trans retinal. This is reduced to all-trans retinol. In the RPE the all-trans retinal is esterifed to fatty acids. The ester is converted by retinoid isomerase to 11-cis retinol. Interphotoreceptor binding protein then returns the 11-cis retinal to the photoreceptor, where it recombines with opsin to form dark adapted RhO (Travis *et al.,* 2007).

1.5 Retinitis Pigmentosa

Retinitis pigmentosa is a broad group of disorders affecting the retina. The main subgroup of diseases is called **retinitis pigmentosa** (RP). For over 140 years RP has been the name given to a group of inherited progressive retinal disorders with a combined incidence of approximately 1/5000 (Rivolta *et al.,* 2002). RP is a highly heterogeneous group of diseases which is characterised by retinal pigment deposits in the midperiphery of the retina (Figure 1.6). Patients with a retinal dysfunction are usually diagnosed with RP based on night blindness which occurs early in life, this will be followed by tunnel vision later in life. RP is inherited by several modes, including autosomal dominant, autosomal recessive, X-linked, digenic and mitochondrial.

1.5.1 Autosomal dominant RP

This is usually the mildest form of RP, with many of the cases not beginning until the later stages of life. 15-20% of all cases of RP are autosomal dominant (adRP) (Inglehearn, 1998). This form of RP is caused by at least 12 genes, including *PRPF31* and *RP1*. Fourteen loci have been mapped for adRP, with more genes to be cloned. The first gene cloned for RP was Rhodopsin (*RHO*) identified in 1990 (Dryja *et al*., 1990). This gene is also the main causative gene for adRP.

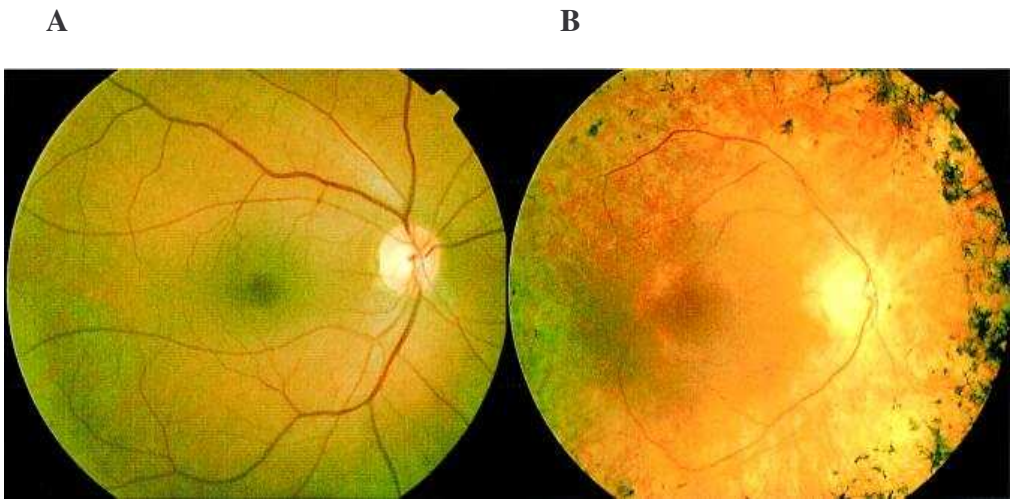


Figure 1.6 Fundus of healthy and RP affected retina. Healthy human retina (A) and retinitis pigmentosa patient retina (B) as seen through the ophthalmoscope. RP retina shows the typical bone spicule pigmentary changes and attenuation of the blood vessels (Farrar *et al.,* 2002).

1.5.2 X-linked RP

X chromosome-linked RP (XLRP), which always presents with severe symptoms, early onset and rapid deterioration, accounts for 6-17% of familial RP cases (Jin *et al.,* 2006). So far five genetic loci have been mapped for X-linked RP, including *RPGR, RP2, RP6, RP23* and *RP26.* Mutations in the *RPGR* gene are estimated to cause 15% to 20% of all cases of RP, higher than any other single RP locus (Bader *et al.,* 2003).

1.5.3 Digenic RP

Digenic inheritance is the interplay of mutations of two different genes resulting in a heritable eye disorder. In this disease mutations in *ROM1* and *RDS* gene combined give a digenic genetic phenotype. Affected individuals are heterozygous for both mutations (Dryja *et al.,* 1997).

1.5.4 Autosomal Recessive RP

In contrast to adRP, autosomal recessive RP (arRP) is typically reported in younger patients and severity is much greater. There are 25 associated arRP loci and for 22 the causative mutation has been identified; (see table) most of these mutations are responsible for a small percentage of cases (1–5%) (Abd-El Aziz *et al.,* 2008). Autosomal recessive RP (arRP) is the most common form of RP worldwide and accounts for 39% of cases in Spain (Abd-El Aziz *et al.,* 2005). Several of the arRP genes are also implicated in other retinal disorders. One such gene causing arRP is *CRB1*; mutations in this gene also cause Leber congenital amaurosis (LCA), which is another clinically heterogeneous retinal disorder.

Table 1.1 Genes and loci for arRP:

Gene	Function	Location	Locus
1. ABCA4	May play a role in photoresponse.	1p21-q32.1	RP19
2. CERKL1	Role in cellular apoptosis and survival	2q31-q33	RP26
3. CNGA1	Involved in Phototransduction pathway	4p14-q13	CNGA1
4. CNGB1	Visual transduction cascade	16q13	_____
5. CRB1	Photoreceptor morphogenesis in the retina Maintains cell polarity and adhesion	1p31-q32.1	RP12
6. LRAT	Catalyzes the esterification of all-trans retinol into all-trans-retinyl ester	4q13	_____
7. MERTK	Associated with disruption of the retinal pigment epithelium (RPE phagocytosis pathway	2q14.1	_____
8. NR2E3	Transcription factor that may have a role i the regulation of signaling pathways intrinsic to the photoreceptor cell function	15q23	_____
9. PDE6A	Phototransduction enzyme	5q31.2-q34	_____
10. PDE6B	Phototransduction enzyme	4p14-q13	_____
11. PRCD	Cytoplasmic protein	17q25.1	_____
12. RGR	Retinoid metabolism	10q23	_____
13. RPE65	Retinoid metabolism	1p31.2	RP20
14. RLBP1	Retinoid metabolism	15q26.1	_____
15. RHO	Visual transduction cascade	3q21-24	_____
16. RP1	Protein transport and disc formation	8q12.1	_____
17. SAG	Visual transduction cascade	2q37.1	_____
18. **EYS**		6q14-q21	RP25
19. TPA		8q12.3	TTPA
20. TULP1	Photoreceptor cell transcription factor		RP14
21. USH2A	Retinal development		_____
22.			RP22
23.			RP28
24.			RP29
25.			RP32

1.6 Syndromic RP

1.6.1 Usher Syndrome

Usher syndrome is a condition that involves both hearing and visual difficulties. Usher syndrome accounts for about 17% of all RP cases in the US (Williams 2007). There are three types of Usher syndrome depending on the severity of both hearing and vision loss. Currently, 12 loci have been found to cause Usher syndrome (Hilgert *et al.,* 2009).

1.6.2 Bardet Biedl syndrome (BBS)

This is a complex disorder which affects children in early childhood. It is also associated with RP. Typical characteristics include obesity, mental retardation, renal abnormalities along with retinal degenerations (Katsanis *et al.,* 2001). It has been found to be linked to 12 loci.

1.7 Other retinal degenerations

1.7.1 Cone-rod dystrophy

Cone rod dystrophies (CRDs) with a prevalence 1/40,000 are inherited retinal dystrophies that belong to the group of pigmentary retinopathies (Hamel, 2007). CRDs can be inherited by autosomal dominant or autosomal recessive means. They are characterised by the primary loss in cone receptors followed by degeneration in the rod photoreceptors. The course of cone rod blindness is normally more severe

and rapid than that of RP, this leads to younger cases of blindness. *RIM1* is one such gene responsible for type 7, CORD7 which is a subtype of CRDs (Johnson *et al.,* 2003).

1.7.2 Leber congenital amaurosis

Leber congenital amaurosis (LCA) causes blindness or severe visual impairment at or within a few months of birth (Wang *et al.,* 2005). It is a highly heterogeneous disorder with several genes implicated; including the recently identified LCA5 gene (denHollander *et al.,* 2007). So far 8 genes and 3 loci have been associated with autosomal recessive LCA. The genes identified are all involved in a variety of physiological pathways, making clinical predictions difficult.

1.7.3 Stargardt's disease

This is also known as fundus flavimaculatus. In this condition there is atrophy of the macula with the appearance of yellow flecks. Mutations in the *ABCA4* gene encoding a photoreceptor-specific, ATP-binding cassette transporter are responsible for the disease (Reinhard *et al.,* 2007).

1.8 Human genome project

Molecular genetics has contributed immensely to our knowledge and understanding of inherited ocular genetics disease. Preceding the development of recombinant DNA technology, the scientific field was limited to phenotypic classifications based on biochemistry. Each human genome contains between 20,000 and 25,000 genes (Abdellah *et al.,* 2004). The human genome project (HGP) was developed to identify all of these genes, along with determining the sequences of the 3 billion base pairs of DNA and storing this data (Venter *et al.,* 2001). The publication of the full genome was completed in April 2003 (Collins et al., 2003). The HGP, the database collections and the advances for disease identification are immense. The powerful combination of the information generated by the Human Genome Project and technical advances such as microarrays enables attempts to identify genes responsible for inherited disorders more possible than ever before. Starting with even modest pedigrees of only a few individuals, or even single individuals, it is now possible to identify the gene(s) involved in the inherited disease.

1.9 Disease gene identification

There is no single pathway to guarantee success in identifying a causative gene. Disease causing genes may be identified by two methods; whole genome candidate gene approach and by linkage analysis. Whole genome candidate gene analysis involves selecting possible causative genes across the genome. By applying this approach, it is possible to identify genes and gene variants that have not previously

been associated with a particular disease. Disease causing genes may, on one hand, be identified based on functional information about the disease and/or on genetic map information obtained by linkage analysis. A drawback to this is that it can only be used on known or previously identified genes, however linkage analysis is the most common and conventional way to identify a disease causing gene.

1.9.1 Linkage analysis

Linkage disequilibrium occurs where alleles at two or more loci appear together in the same individual more often than would be expected by chance (LaFramboise *et al.,* 2009) In this approach, the aim is to find out the approximate location of the gene relative to another DNA sequence called a genetic marker, which already has a known location on the chromosome. The principals of linkage are simple, in that we inherit maternal and paternal alleles, in which there are sequence variants. When this involves a disease allele, it should be easy to distinguish from the normal allele by searching for incidents of disease in the family tree.

1.9.2 Genetic markers

Genetic markers are DNA sequences that show a physical location on a chromosome and whose inheritance can be followed. A marker can be a gene, or it can be some section of DNA with no known function (Smith and Ebrahim 2004). Because DNA segments that lie near each other on a chromosome tend to be inherited together,

markers are often used as indirect ways of tracking the inheritance pattern of a gene that has not yet been identified, but whose approximate location is known.

Genetic markers first became of use in the 1980's, when it was discovered that DNA polymorphisms can be used as a marker system due to their prevalence across the human genome. Commonly used genetic markers include RFLPs (restriction fragment length polymorphisms), VNTRs (variable number tandem repeats) and SNPs (single nucleotide polymorphisms).

1.9.3 Single nucleotide polymorphisms (SNPs)

SNPs are DNA sequence variations that occur when a single nucleotide is altered. For a variation to be considered a SNP, it must occur in at least 1% of the population. SNPs compose about 90% of all human genetic variation occurring every 100 to 300 bases along the 3-billion-base human genome (Chakravarti, 2001). SNPs may occur in both coding and non-coding regions of the genome and these changes can have a major effect on a person's response to a disease, toxins, and chemicals (McCarthy and Hilfiker 2000). Until recently, the majority of gene-mapping studies have focused primarily on SNPs in human disorders. However another type of DNA variation has been recognised as an important source of genetic variation, these are called copy number variants (CNVs).

1.9.4 Copy Number Variation (CNV)

Initially it was assumed that the majority of genetic variation in the genome was due to SNPs. However over the past number of years there is a growing interest in the study of structural genetic variation. Copy number variants (CNVs) are now thought to be a prevalent form of common genetic variation and represent a large proportion of genetic variability in human populations (Ionita-Laza *et al.,* 2008). A CNV is described as a DNA segment that is 1 Kb or larger and present at variable copy number in comparison with a reference genome (Redon *et al,* 2006). Deletions, insertions, duplications and complex multi-site variants are collectively called CNVs or copy number polymorphisms (CNPs) (Redon et al, 2006). In comparison to the well-developed resources for SNP analysis, CNV-association studies are still in the early stages, however the development of statistical tools and databases is anticipated in the coming years.

Since the completion of the mapping of the human genome, an immense amount of data has been collected. Vast amounts of data associated with disease causing genes, novel SNPs and other information associated with both coding and non coding regions of genes has been gathered. All of this data has had to be collected, stored and analysed in depth. This has encouraged the development of databases and has promoted the development of bioinformatics.

1.9.5 Databases and Bioinformatics

Over the past decade or so, there has been a great increase of genetic databases and biobanks, which promise to increase the understanding the way our genes interact with the environment. These databases are a crucial resource for identifying disease causing genes. Technologies such as genome-sequencing, microarrays, proteomics and structural genomics have provided vital information for many living organisms, and researchers are now focusing on how the individual components fit together to build systems (Collins *et al*., 2003). There is an ongoing and growing need to collect, store and interpret all of this information in ways that allow its efficient retrieval and utilization. The European Bioinformatics Institute (EMBL-EBI), which is part of the European Molecular Biology Laboratory (EMBL), is one of the few places in the world that has the resources and expertise to fulfill this important task. Its aim is to develop a software system that produces and maintains automatic annotation on eukaryotic genomes. The use of programs such as Ensembl, allows us to retrieve data on the genomes of numerous species. It gives the genetic sequence and structure of thousands of genes. Access to such information allows for the interpretation and prediction of gene makeup.

1.10 Mutation Identification Techniques

Mutation detection techniques are based on testing DNA samples for changes in or near the relevant gene. Currently the most common means is through automated fluorescence sequencing, as it is quick and cheaper than other methods. Alternative methods include denaturing high performance liquid chromatography (dHPLC), single strand confirmation polymorphism (SSCP) and quantitative PCR. Direct sequencing has surpassed these methods due to its efficiency and accuracy in mutation detection (Han *et al.,* 2004). Methods such as SSCP are based on heteroduplex analysis, which uses electrophoretic mobility in gels to identify abnormalities.

1.10.1 Identifying and confirming disease causing mutations

A very important factor when determining a disease gene is to confirm that the pathogenic change segregates with the disease phenotype within a pedigree. This change must be tested in a control panel of unrelated ethnically matching control individuals. A mutation can also be tested using a restriction digest, which is an enzymatic technique used to cleave DNA at specific sites yielding fragments of specific sizes (Botstein et al., 1980). A mutation can either create or delete specific restriction fragment sites. Other methods such as direct sequencing analysis and SSCP can also be used to test this segregation. Once a gene has been identified as being implicated in a particular disease, it is crucial to determine what it does under normal conditions, and

how it changes when it possesses a mutation. Sometimes there is a wealth of information available in relation to the gene and the genes function, however, occasionally the identified gene is novel and its structure and expression have to be established. Characterisation of a novel gene involves determining the full coding and non coding regions, along with 5' and 3' untranslated regions (UTRs).

1.11 Further Analysis of a Novel Gene

1.11.1 Protein Structure Identification

Protein structure is also an important feature to identify the proposed function of the gene implicated in a disorder. Domains identified in the protein are compared with domains of characterised proteins to predict the function of this new protein. A protein domain is the part of the protein structure that can evolve, exist and function independently of the rest of the protein (Jain *et al.,* 2009). Each domain in a protein forms a three-dimensional structure and can be often independently folded. It is common for proteins to have several structural domains and vary in length from between 25 amino acids up to 500 amino acids and are evolutionary conserved. This conservation is essential when identifying the structure and function of a novel protein. If homology is identified between the protein structure of the novel protein and a previously characterised protein, then the function of this new gene can be speculated.

1.11.2 Functional analysis of a novel protein

Antibody production is the primary step in the investigation of the functional study of a gene of interest. Antibodies are tools to demonstrate both the presence and the sub cellular localisation of a protein through the antigenic epitopes in the polypeptide. Studies to investigate the function of a novel gene include studies of the normal gene in living cells and tissue. Immunohistochemical and immunocytochemical techniques utilise the antibody that targets the specific protein through the use of other labeled antibodies or markers (Ramos-vara, 2005). Other experiments involve examining the same normal gene in animal models such as mice, rats or fruit-flies. Similar studies to those performed for the normal gene must also be executed on the mutated copy of the gene, to investigate how the changes alter the function of the protein that the gene encodes.

1.12 RP25 Locus and refinement of Locus

1.12.1 Background

The *RP25* gene is the 26th gene to be identified for arRP (Abd El-Aziz *et al.,* 2008). The locus for RP25 was originally established through homozygosity mapping by targeting functional candidate genes. Four Spanish families with arRP were initially

linked to a region on chromosome 6q between microsatellite markers D6S257 and D6S1644 (Ruiz *et al.*, 1998). The linkage was further refined when another three families were linked to the interval (Figure 1.7).The region flanked by microsatellite markers spans ~ 34 Mb in size and has approximately 110 genes mapped in between these markers. The clinical data for five of these families is depicted in table 1.2. Candidate genes were initially selected based on their known function of expression patterns and other genetic data available on the genes. Several genes such as *GlcAT-S* (Marcos *et al.,* 2002) and *RAB23* (Marcos *et al.,* 2003) were selected as good candidate genes based on their expression profiles. *GlcAT-S* was selected due to its high expression level in the retina and *RAB23* was selected because proteins belonging to the Rab family had previously been implicated in retinal degenerations. Several more genes were also selected and excluded as causative genes for RP25, these included *KHDRBS2, PTP4A1, KIAA1411, OGFRL1* and *FAM46A* (Abd El-Aziz *et al.,* 2005) (Barragan *et al.,* 2008). Further information on candidate gene selection is found in section 3.1. A strategy was also undertaken to divide the interval into 4 regions, A, B, C and D (Section 3.1) based on the linkage of another family of Pakistani origin which mapped north of the RP25 interval (Figure 1.8), three Chinese families were also mapped to the RP25 interval (Abd El- Aziz *et al.,* 2007).

1.12.2 Families of other ethic origin linked to RP25

1.12.2.1 Pakistani family

Further to the linkage of Spanish families, in 1999 Khaliq and colleagues mapped a Pakistani family to the same interval. In the study 20 members of a consanguineous family who displayed features of autosomal recessive retinitis pigmentosa were used. Such features included night blindness commencing at approximately 25 years and deterioration of central vision at ~30 years of age. All 12 affected members of the family were analysed. Genomic DNA from each individual was collected and they were genotyped for microsatellite markers for all the known arRP loci. When linkage was discovered on 6q, further polymorphic markers from this region were analysed to determine whether analysis of recombinant individuals within the family would permit further refinement of the published disease interval (Figure 1.8).

1.11.2.2 Chinese families linked to the RP25 interval

Three non-consanguineous Chinese families were also linked to the *RP25* interval and were linked using a 10K GeneChip. This was undertaken to detect regions of shared haplotypes between the affected members in each family. The three Chinese families were annotated RP10, RP112 and RP116 respectively (Figure 1.8).

Table 1.2 Clinical Characteristics of RP25 patients:

Family	Age	Age of onset	Symptoms	Visual fields	Fundus	ERG
RP5	56	15	N,VFC,VAR	CC IN BE	BS & POD	NR BE
RP73	36	20	N,VFC,VAR	CC in BE	BS & POD	NR BE
RP167	57	38	N,VFC,VAR	CC in BE	BS & POD	NR BE
RP214	37	30	N,VFC	CC in BE	BS & POD	NR BE
RP299	38	15	N,VFC	CC in BE	BS & POD	NR BE

N: Nyctalopia
VFC: Visual field constriction
VAR: Visual acuity reduction
CC: Concentric constriction
BE: Both eyes
BS: Bone spicules
POD: Pale optic disc
NR: Not recordable

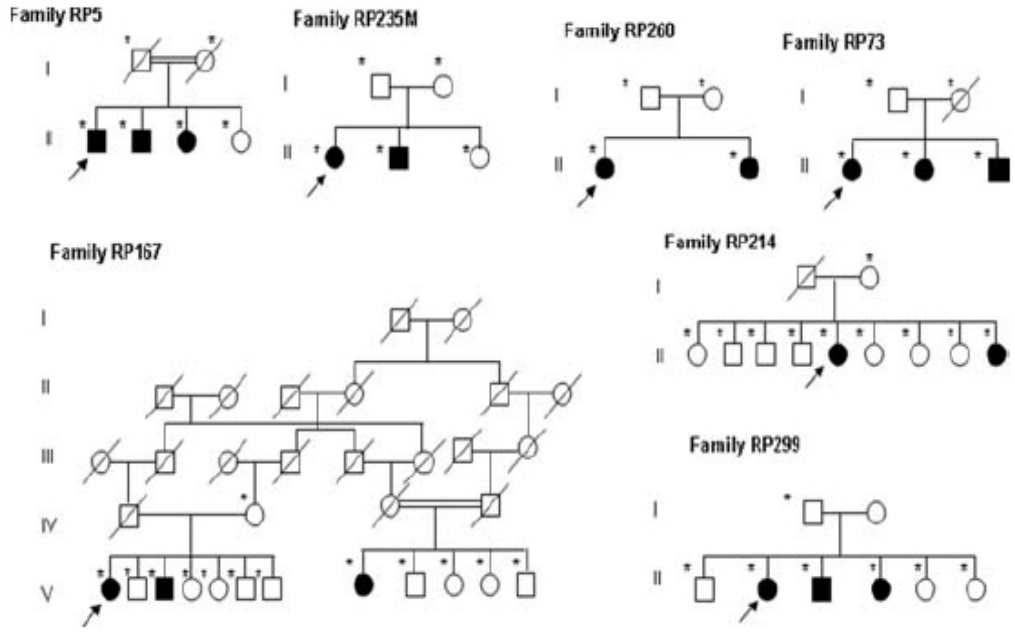


Figure 1.7 Seven Spanish families linked to the RP25 interval. The arrows denote which members of the family were first affected with RP. The families which are consanguineous are RP5 and RP167, and nonconsanguineous are RP235, RP260, RP73, RP214 and RP299. The arrows denote which members of the families were first identified as having arRP symptoms. The asterisk corresponds to the DNA of members we had available for use in the screening process.

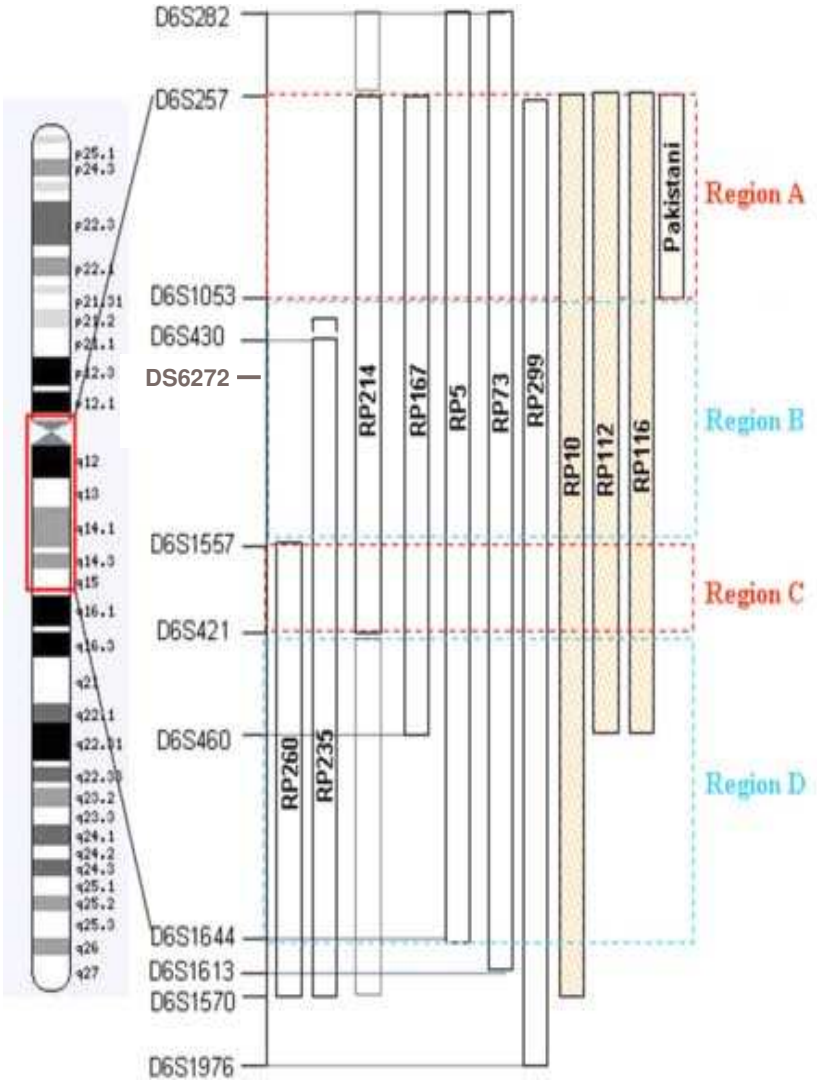


Figure 1.8 Overview of the arRP Spanish affected families, along with the Pakistani and Chinese families whom have been linked to the RP25 interval. The Pakistani family has been linked to an interval of 8.61 Mb spanning from microsatellite marker D6S257 to D6S1053. The combined Chinese families span an interval of 16.1 Mb, from microsatellite marker D6S257 toD6S421.

1.12.3 Chromosome 6q an important interval for retinal degenerations

Another important aspect of this study is that chromosome 6q is a central genetic interval for several other retinal degenerations. These include, cone rod dystrophy (CORD7), leber congenital amarousis (LCA5), dominant drusen (DD) and several others (Figure 1.9). The occurrence of multiple loci so closely mapped in this chromosomal region could indicate the presence of a number of retinal genes in continuum, since the phenotype of all these retinal dystrophies, their ophthalmologic appearance, age at onset, and the extent and pattern of visual loss are varied (Dharmaraj *et al.,* 2000). The other theory was that one single large gene in this region would be the cause of numerous retinal degenerations, as in the case of *ABCR.* Mutations in this gene cause numerous retinal degenerations including arRP (Martinez-Mir *et al*., 1998) cone-rod dystrophy (Cremers *et al.,* 1998) and recessive Stargardt disease (Kaplan *et al*., 1993).

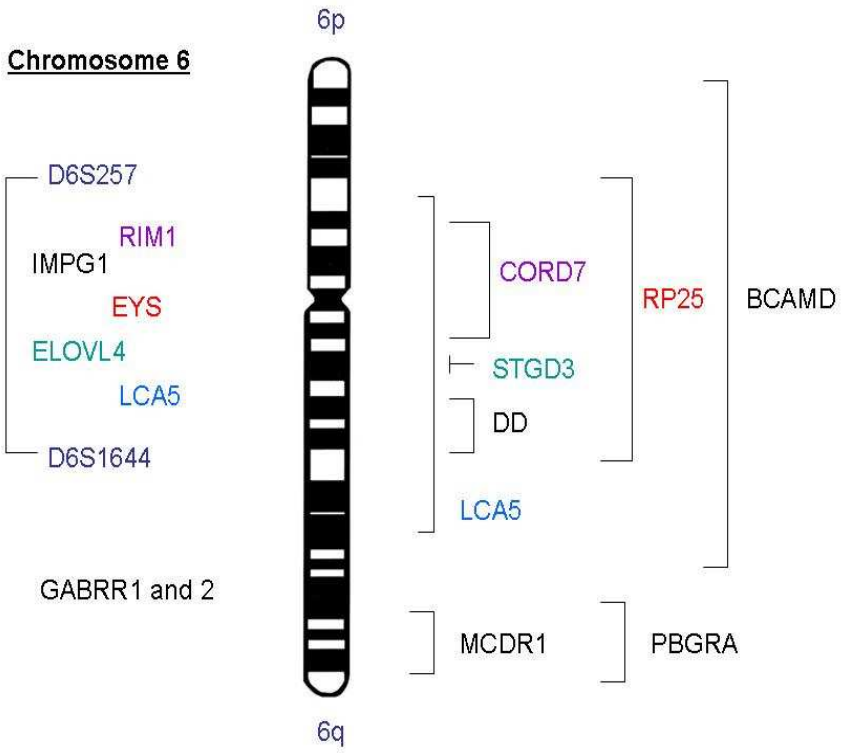


Figure 1.9 Depiction of chromosome 6 and of the corresponding retinal disease associated genes and loci. Note the location of the RP25 locus (red) between microsatellite markers D6S257 and D6S1644. Disorders include Cone-rod dystrophy (CORD7), Stargardt disease (STGD3), Dominant drusen (DD), Leber Congenital Amaroursis (LCA5), Benign concentriannular macular dystrophy (BCAMD), North Carolina Macualr degeneration (MCDR1). The colours of the genes correspond to the same colour disorder.

1.13 Aim of Project

The aims of this project include systematically selecting and sequencing candidate *RP25* genes in the RP25 interval. Alternative methods to identify mutations including CNV analysis is also incorporated into the study. Further aims of the project include recruitment of more RP25 affected families to narrow down the RP25 interval. The overall goal of the project is to identify the RP25 gene and to determine the possible function of this disease causing gene.

Chapter 2

Materials and Methods

2.1 PCR

The polymerase chain reaction (PCR) is a biochemical and molecular biological technique. It is used to isolate and amplify specific fragments of DNA exponentially. PCR is a relatively recent invention, only created in 1983 by Kary Mullis and is a commonly used technique in biological research. It can be used to clone genomic sequences with great efficiency, directly sequence genomic and mitochondrial DNA, and also to analyse nucleotide sequences (Bartlett and Stiling 2008). The overall purpose of PCR is to generate a large number of copies of a particular sequence; which is crucial in generating enough starting template to enable downstream applications.

2.1.1 The cycling reactions

There are three major steps in PCR, which are repeated for 30 or 40 cycles. This is done on an automated cycler, which can heat and cool the reaction mixture in a very short time. These include

1: Denaturation at 94°C during which, the double stranded DNA melts open to single stranded DNA, all enzymatic reactions stop. Initial denaturation is at 94°C for 3 min, 35 cycles of denaturation at 94°C for 30 seconds.

2: Annealing at 55-68°C. In this step, the primers are moved around; hydrogen bonds are constantly formed and broken between the single stranded primer and the single

stranded template. This is known as the annealing stage and occurs at 55°-68°C for 30 seconds (depending on the required annealing temperature).

3: Extension at 72°C: The temperature in this step is ideal for the polymerase to work. The bases (complementary to the template) are coupled to the primer on the 3' side (the polymerase adds dNTPs from 5 ' to 3 ', reading the template from 3 ' to 5 ' side; bases are added complementary to the template). Extension at 72°C for 30 seconds (time depends on the size of the amplicon), final extension step of 72°C for 5 min.

2.1.2 Taq polymerase

The enzyme used in the PCR reaction is called *Taq* polymerase. It is a highly thermostable DNA polymerase from the thermophilic bacterium *Thermus aquaticus*. The enzyme catalyzes template-dependent polymerization of nucleotides into duplex DNA in the 5 '=>3 ' direction. *Taq* DNA Polymerase exhibits deoxynucleotidyl transferase activity, which frequently results in the addition of extra adenines at the 3 '-end of PCR products (Peake, 1989).

2.1.3 Parameters for PCR

All PCR reactions carried out in this study were performed in 25µl reaction volume containing 1X buffer (10X NH4 buffer, Bioline, UK), $MgCl_2$ (1.5, 2.5 and 3.5 mM, Bioline, UK), 0.2 mM deoxynucleoside triphosphate (dNTPs: dATP, dCTP, dGTP, dTTP, Promega, UK), 10 pmoles of each primer, 0.5 unit of *taq* polymerase (Bioline,

UK) and ~100ng of DNA. Amplification reactions were performed using a Thermal Cycler equipped with heated lids. The amplification conditions are as above.

2.1.4 Special parameters

For GC rich areas, buffer 3 (containing 22.5 mM $MgCl_2$; supplied by Boehringer Manheim GmbH-Germany) together with dimethyl sulphoxide DMSO 5% or ready mix (AB-0975; ABgene, UK) were used.

2.1.5 Annealing temperature and primer design

Both primer length and annealing temperature are fundamental components when preparing a successful PCR reaction. The melting temperature of a duplex reaction increases both with its length, and with increasing (G+C) content. There is a simple formula to calculate the melting temperature (T_m): Tm=4(G+C) +2(A+T)ºC

The annealing temperature was dependent on length and nucleotide composition of the primers. The length of primers would generally be between 18 and 27 nucleotides, the difference between T_m of both primers should be no greater than 4°C. There should also be a random base distribution, 40-60% GC within the primer designed. Along with these rules, it is also necessary to ensure low complementary of each primer to each other, to avoid the formation of primer dimmers. In this project, the software program Primer 3 Output (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) was used to design primers, commercially synthesised primers were purchased from Sigma

Aldrich (http://www.sigmaaldrich.com/united-kingdom.html). See appendix for table of primers used in this thesis.

2.2 Agarose Gel Electrophoresis

Gel electrophoresis is a method used to separate DNA and RNA molecules based on size. Separation was performed by an electrical current supplied to the gel tank. To prepare the gel for the separation a substance called agarose was used. Agarose comes from a family of polysaccharides called agars that are obtained from algae such as seaweed and structurally it is a galactan (galactose polymer) (Yaphe 1984). It is purified from agar and different purities are available with varying melting temperatures.

Depending on the concentration of the gel and the purpose of its use (whether RNA or DNA analysis) a quantity of agarose was weighed in a clean conical flask. A relative amount of 1X TAE buffer was added to the agarose powder and was mixed gently. The mixture was heated in a microwave to allow the agarose to dissolve. Once the mixture had cooled to approximately 50°C, ethidium bromide was added. This is an intercalating intercalating agent which allows staining of nucleic acids (Biggiogera and Biggiogera 1989). A sealed loading tray with a set number of combs was prepared. The cooled agarose gel was poured into this tray and left to set. Once the gel was polymerised, the combs and seals were removed and the gel was placed in the electrophoresis tank (Gel electrophoresis apparatus GNA-200, Pharmacia, Sweden) which contains 1X TAE buffer. An electrical current was run through the tank from

positive to negative. In the tank, the DNA samples mixed with the loading dye were added to each lane of approximately 10µl. A suitable DNA ladder was also loaded to detect the size of each band. Electrophoresis was performed at 100-120 V for 30 min, gels were analysed using a biodoc-it imaging system (UVP).

2.3 Purification of PCR products

2.3.1 Montage cleanup purification

The Montage SEQ_{96} Sequencing Reaction Cleanup Kit provides the filtration plate and the solution necessary to remove contaminating salts and unincorporated dye terminators from DNA sequencing reactions. It is based on the incorporation of a vacuum pump and the filtration plates. The vacuum-based size separation platform eliminates the need for gel filtration or bind-wash-elute columns that require centrifugation. The protocol can be altered slightly depending on the concentration of the PCR products. Initially 80 µl dH_2O was added to a 20 µl of the PCR products. The whole volume was transferred to the filtration plate and attached to the vacuum pump; which was set to vacuum for 10-15 min. 25 µl of dH_2O was then added to the plate and was vacuumed again for 2-3 min. Finally 20 µl of dH_2O was added to the plate which was left to shake on the vortex at 1000 rpm for 10 min to elute the DNA. The cleaned products were then collected; which were directly used for sequencing or were stored in the -20°C freezer for future use.

2.3.2 ExoSAP IT

ExoSAP-IT is another efficient clean-up system; it removes contaminants while utilising two hydrolytic enzymes, Exonuclease I (SAP, QAmersham LifeScience, Buckinghamshire, UK) and Shrimp Alkaline Phosphatase (United States Biochemicals, Ohio, USA) mixed together in a specially formulated buffer, to remove unwanted dNTPs and primers from PCR products. Exonuclease I removes residual single-stranded primers and any extraneous single-stranded DNA produced in the PCR; while Shrimp Alkaline Phosphatase removes the remaining dNTPs from the PCR mixture

To set up the exosap reaction, 1 µl of the amplified PCR product was mixed with 16.5µl of dH_2O along with the enzyme mix exoSAP IT. This combination was incubated at 37°C for 15 min followed by 80°C for another 15 min to deactivate the enzyme. The product was used for the following sequencing reaction.

2.4 Automated DNA sequencing

Automated DNA sequencing uses a fluorescent dye to label nucleotides rather than using a radioactive isotope. Fluorescent dye can be used without any special requirements or reagents that hazardous to use in any way. Rather than using X-ray film to analyses the sequence, a laser is used to stimulate the florescent dye. The emissions are collected and the wavelengths are determined. In this study, the ABI 3100 sequencer was used to analyse sequences. For ExoSAP the following protocol was used: the product (18µl) was mixed with 0.5µl Big Dye and 0.5µl of forward or

reverse sequence specific primers and 2.5µl sequencing buffer (Applied Biosystems). Cycle sequencing reactions were performed in the thermal Cycler and the temperature cycling profile consisted of initial denaturation at 96°C for 3 min followed by 25 cycles of denaturation at 95°C for 10 seconds, annealing at 50°C for 5 seconds and then 4 min at 60°C, and finally a further extension at 4°C for 2 min. After completion of the cycle sequencing reaction the DNA was precipitated using sephadex clean up protocol. Sephadex powder (Sigma, UK) was put into a Millipore multiscreen plate. 350µl of dH_2O was added to each well. A collection plate was placed underneath the multiscreen plate, and the plate was spun for 5 min at 900 rpm. 100 µl of dH_2O was then added to each well. The plate was again spun for 5 min at 900 rpm. Finally the DNA samples were added to the plate and the final spin was run for 5 min at 900 rpm. 10µl of dH_2O was added to each well of the collection plate and the DNA samples were read by the automated DNA sequencer. Sequences were analysed using Seqman software package.

2.5 Restriction Enzyme Digest Analysis

Restriction digests were performed on PCR products to identify if a mutation is present in a selected patient or panel of control DNA samples. For PCR product digests, a total volume of 20µl containing 2µl 10X buffer 2, 0.2µl 1X Bovine serum Albumin (BSA) to enhance the enzymatic activity, 10µl PCR product, 0.5-1.0µl of the appropriate enzyme (1 units) and 7.3µl dH_2O was used. The reactions were then incubated for at least 2 hrs or overnight at the specified temperature for the enzyme (usually 37°C). After heat inactivation the product of the restriction digest were visualized by agarose

gel electrophoresis (see previously section 2.2). Enzymes were obtained from New England BioLabs (UK) or Promega, UK.

2.6 SMART RACE cDNA Amplification

The SMART RACE cDNA Amplification Kit (Clontech, France) provides a method for performing both 5`- and 3`- rapid amplification of cDNA ends (RACE). This kit integrates the Marthon cDNA Amplification Kit with the SMART (Switching Mechanism At 5` end of RNA Transcript) cDNA synthesis technology. This powerful combination allows isolation of the complete 5` sequence of the target transcript. The SMART RACE Kit includes recent advances in the PCR technology that both increases the sensitivity and reduces the background of the RACE reactions. SMART technology provides a mechanism for generating full-length cDNAs in reverse transcription reactions (Zhu *et al,* 2001)

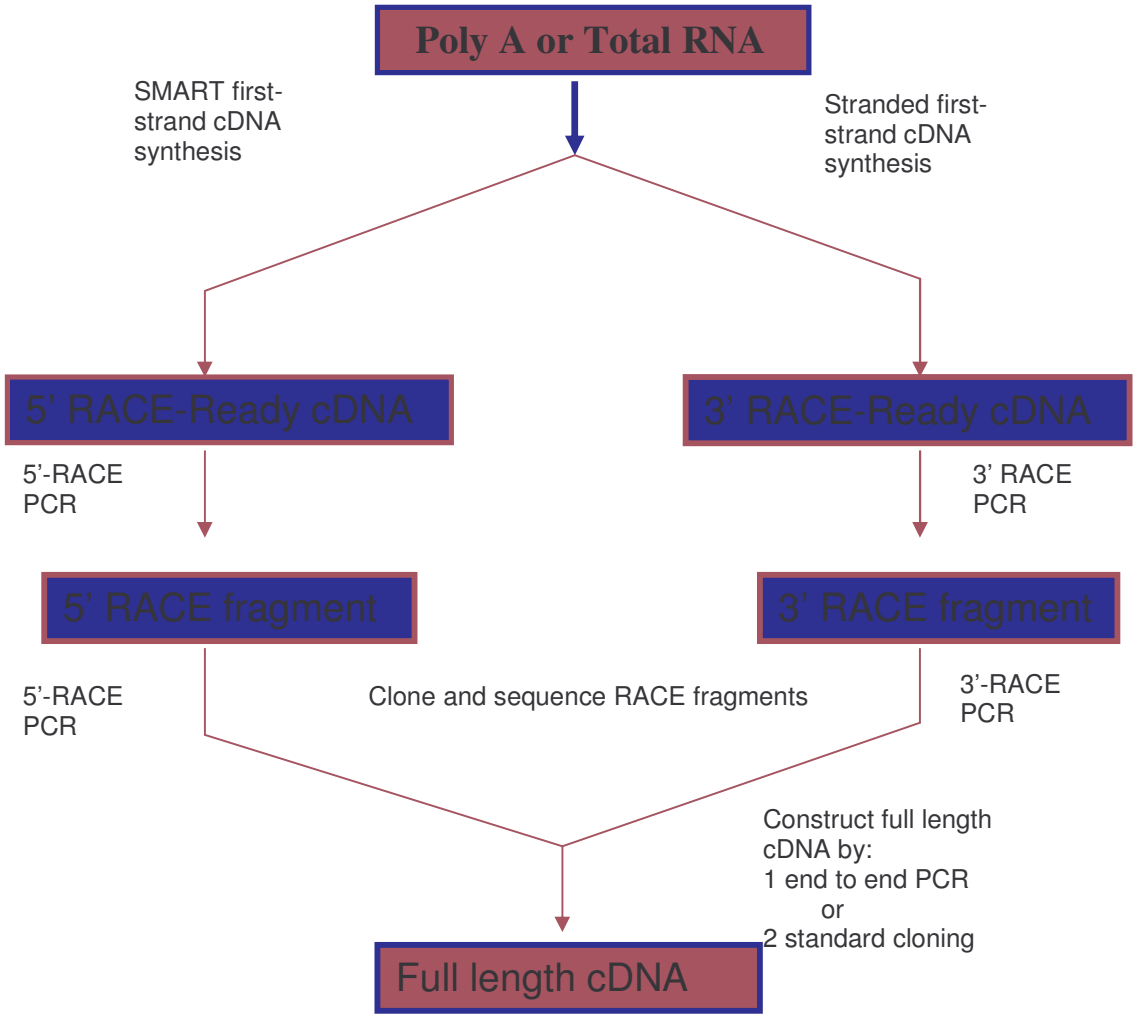


Figure 2.1 Overview of RACE protocol. Adapted from Roche website.

2.6.1 Primer Design

Gene specific primers (GSPs) are for the 5` and 3` RACE reactions. These must be between 23 and 28 base pairs in length, must have a GC content over 50%, and a T_m > 70°C.

First-strand cDNA synthesis

To prepare cDNA for 5` and 3` RACE , 1- 3µl of the RNA sample was added to a cDNA specific primer (CDS primer A) and 1µl SMART II A oligonucleotide. Sterile H_2O was added to each tube to make up a volume of 5µl for each reaction. Contents were spun briefly in a centrifuge, and incubated at 70° for 2 min. Tubes were cooled on ice for 2 min, and again spun briefly. The following were added to each tube: 2µl of 5X first-strand buffer, 1µl DTT (20mM), 1µl dNTPs (10mM), 1µl PowerScript Reverse Transcritase to make a total volume of 10 µl. The contents were again spun down and incubated at 42°C for 1.5 hr in a thermal cycler. The products were diluted with 250 µl of tricine-EDTA buffer, the samples were heated at 72°C for 7 min. The samples were stored at -20°C for up to 3 months.

Positive Control RACE experiment

Prior to performing RACE with our own template, a positive control RACE experiment was performed using the control Human Placental Total RNA provided with the kit. A master mix was prepared. For each 25 µl PCR reaction, the following reagents were added:17.25 µl PCR-grade water, 2.5 µl 10x Advantage 2 PCR Buffer, 0.5l µl dNTP Mix, 0.5l µl 50X Advantage 2 Polymerase Mix to make to a total of 20.75 µl.

To prepare for a PCR the below table outlines the order and the components of each tube

Table 2.1 PCR preparation for the 5`-RACE and 3`-RACE reactions:

Component	5` RACE control	3`RACE control	Internal Control (5`-cDNA)	Internal Control (3`-cDNA)
Control 5` RACE Ready cDNA	1.25 µl	_____	1.25 µl	_____
Control 3`-RACE Ready cDNA	_____	1.25 µl	_____	1.25 µl
5`-RACE TFR Primer (10 M)	0.5 µl	_____	0.5 µl	0.5 µl
3`-RACE TFR Primer (10 M)	_____	0.5 µl	0.5 µl	0.5 µl
UPM (10X)	2.5 µl	2.5 µl	_____	_____
Master Mix	20.75 µl	20.75 µl	20.75 µl	20.75 µl
Final Volume	25 µl	25 µl	25 µl	25 µl

Footnote

TFR-gene specific primer for transferrin

UPM: Universal Primer Mix.

Thermal cycling was undertaken using the following touchdown PCR program. The following reaction was set up: 5 cycles at 94°C for 30 seconds, then at 72°C for 3 min. This was followed by another 5 cycles at 94°C 30 seconds, followed by 70°C for 30 seconds, and at 72°C for 3 min. The final set of cycles consisted of 27 cycles at 94°C for 30 seconds, 68°C for 30 seconds and 72°C for 3 min. Finally, it was possible to generate the 5`-RACE and 3`-RACE PCR fragments. PCR Master mixes were made up for each reaction: 17.25 µl PCR-grade water, 2.5 µl, 10x Advantage 2 PCR Buffer, 0.5 µl dNTP Mix, 0.5 µl 50X Advantage 2 Polymerase Mix to make a total volume of

20.75 µl. These were mixed well by vortexing. The following tables depicts the preparation for the 5`-RACE and 3`-RACE.

Table 2.2 Preparation for the 5`-RACE and 3`-RACE reactions.

Component	1: 5`/3`-RACE sample	2: 5`/3`-TFR (+ Control)	3: GSP 1 + 2 (+ Control)	4: UPM only (- Control)	5: GSP1 only (- Control)
5`/3`-RACE-Ready cDNA	1.25 µl	1.25 µl	1.25 µl	1.25 µl	1.25 µl
UPM (10X)	2.5 µl	2.5 µl	_____	2.5 µl	_____
GSP1 (10 M)	0.5 µl	_____	0.5 µl	_____	0.5 µl
GSP2 (10 M)	_____	_____	0.5 µl	_____	_____
Control 5`/3`-RACE TFR Primer (10 M)	_____	0.5 µl	_____	_____	_____
H_2O	_____	_____	2 µl	0.5 µl	2.5 µl
Master Mix	20.75 µl	20.75 µl	20.75 µl	20.75 µl	20.75 µl
Final Volume	25 µl	25 µl	25 µl	25 µl	25 µl

Footnote:

GSP-gene specific primer

2.6.2 Alternative cDNA synthesis protocol

cDNA was prepared from RNA using SuperScript™ II Reverse Transcriptase (Invitrogen, UK). 1µg of RNA was added to RNase free tube and was made up to 4µl with RNase free water. These tubes were then placed in a water bath at 65°C for 10 min. The tubes were put on ice immediately and were then spun down and a reaction mix made of the following components: 5µl of 5x First Strand buffer, 2µl of 0.1 M

DTT (Dithiothreitol), SuperScript™ II Reverse Transcriptase (Invitrogen, UK), 2 µl of dNTP mix (Promega, UK), 1µl RNasin® Plus RNase Inhibitor (Promega, UK), 1µl of pd (N_6) random hexamers (Amersham Biosciences, UK) and was added and the total volume made up to 16 µl with RNase free water. This was added to the cDNA synthesised and was kept at 37°C for 2 hrs. The samples were then stored at -20°C. 1µl of these samples were then subjected to the Quantitect Probe RT-PCR reaction.

2.7 RT-PCR

The principles of RT-PCR are straight forward and the protocol included three main steps. 1-RNA was reverse transcribed to cDNA using a reverse transcriptase enzyme and primers. This is also known as the reverse transcription step. 2- The dsDNA was denatured at a high temperature so the two strands separated and the primers can bind at a lower temperature and start a new reaction. 3- The DNA extension step involved using thermostable Taq DNA polymerase at 72°C. This is a highly sensitive technique, even detecting very low copy numbers of RNA.

2.8 Immunohistochemistry (IHC)

Immunohistochemistry is the demonstration of antigens within tissue sections by means of specific antibodies and the binding of antigen-antibody is demonstrated through a coloured histochemical reaction visible by light microscopy or fluorochromes with UV (Ramos-Vara 2005). There are numerous immunohistochemical methods that may be used to localise antigens. The selection of a suitable method should be based on

parameters such as the type of specimen under investigation and the degree of sensitivity required.

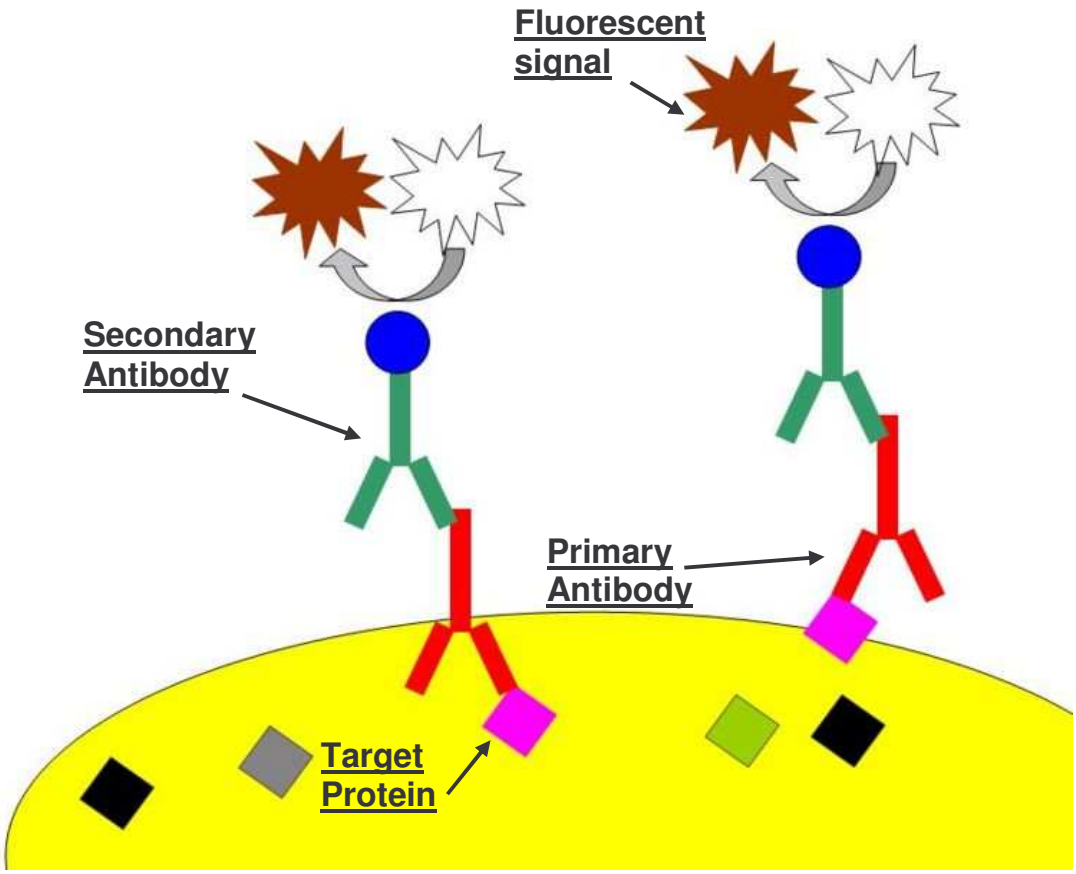


Figure 2.2 Overview of immunohistochemical protocol. The target protein (pink) is targeted using a specific antibody raised against a peptide in this protein. The antibody (red) attaches to the protein of interest. A secondary antibody (green) is added. This is usually fluorescently tagged. This binds to the primary antibody and emits a signal.

2.8.1 Preparation of eye pieces

Porcine eyes were sourced from a local abattoir, where boars or sows were sacrificed in the morning and the eyes extracted within 1 hr of slaughter. The eyes were immediately transported to the laboratory; the retina was extracted and instantly transferred into a fixative solution. On the first day, the pig was sacrificed; the eye was removed with 2-3 mm of the optic nerve. The cornea and lens were carefully removed so as not to cause retinal detachment. The eye was cut into pieces using sterile instruments and is left in 4% paraformaldehyde (PFA) in the fridge overnight. The next day, the eye piece was rinsed with 1xPBS 3 times each for 15 min consistently on a shaker. The piece was then placed in 5% sucrose solution diluted in 1x PBS and left to shake at room temperature for 30'. The eye piece was methodically transferred into a gradient of sucrose diluted solutions. The sucrose gradients were as follows: 2/3 5% sucrose and 1/3 20% sucrose, ½ 5% sucrose and ½ 20% sucrose, 1/3 5% sucrose and 2/3 20% sucrose, all for 30' shaking at room temperature. Finally, the eye piece was left in 20% sucrose at 4°C. The next day, it was placed in 2/3 20% sucrose and 1/3 OCT (freezing medium, Thermos biolabs UK) and was left to shake slowly at room temperature for 30'. The eye piece was then placed in a cylinder of OCT only at room temperature for 30'. Using a forceps the eye piece was orientated in a small plastic cylinder of OCT. A plastic beaker filled with 2-Methlybutane was dipped into a flask of liquid nitrogen. The plastic container with the eye was then placed into the 2-Methlybutane until the OCT freezes completely. The eye piece was then stored at -80°C.

2.8.2 Cryosectioning of eye pieces

The eye pieces were removed from the -80°C freezer and placed inside a cryostat. The cryostat was prepared in the following way, a blade was placed inside the cryostat, and the eye piece embedded in OCT was removed from its plastic holder. This was then placed in fresh OCT and was left to set inside the cold cryostat. Once set, it was placed horizontally from the blade and sections were cut and disposed of. Once the eye piece was clearly visible, the cryostat was set to 10µm and sections were cut one at a time. These were then placed carefully on the surface of superfrost ultra plus slides (Thermo Scientific, UK). The slides were labelled; kept cold until all are collected, and are then stored in the -80°C freezer.

2.8.3 Immunohistochemical Protocol and parameters

IHC experiments have many parameters by which they may success of fail by. One of the major parameters is the production of a good antibody. Another is the quality of tissue being used. The tissue must be relatively fresh and be stored adequately. Tissue is usually stored in PFA, as this is found to be a very good tissue structure preserver. In many instances, an antigen signal many be difficult to obtain. In such cases pre-treatment solutions are required; these include using sodium borohydride, TRIS citrate and citrate acid. For pig retina sections the most successful pre-treatment was sodium borohydride.

2.8.3.1 Pre-treatments

In the event of unmasking an antigen signal, pretreatment solutions such as Sudan black and sodium borohydride (Sigma Aldrich, UK) are often integrated into the immunohistochemical protocol. Sudan black B is a fat soluble dye used for staining neutral triglycerides and lipids on frozen sections. Sodium borohydride is a white powder that is soluble in water. It is recognised as a successful treatment to unmask antigens, particularly in glutaraldehyde fixed tissue by reducing glutaraldehyde linkages (Robinson, 2001). For these experiments the most successful pretreatment was sodium borohydride.

1:Pretreatment with Sodium borohydride

Frozen sections were removed from the -80°C freezer, and were placed in a slide tank of 0.1% sodium borohydride. The sections were left in solution for 30 min, after which they were transferred to another slide tank with of 1% tris buffer saline (TBS). The sections were left to shake for 10 min.

2:Immunolabelling for Cryo-sections (Standard)

A hydrophobic pen was used to outline a small area around the retinal section on the slide and the sections were blocked for one hr at room temperature, in 0.1% normal goat serum (NGS) in 0.1% triton-X with phosphate buffer saline (PBS solution) (PBS-triton) in a humidity chamber. The slides were washed three times in PBS-triton for 10

min each, and incubated with the primary antibody (diluted according to the manufacturer's recommendation or in serial or different dilutions for novel antibodies, in 2% NGS in PBS-triton) for one hr at room temperature or alternatively at 4˚C overnight. The slides were washed three times in PBS-triton for 10 min each. The corresponding secondary antibody was diluted 1:300 in 2% NGS in PBS-triton and the slides incubated at room temperature for an hr in the dark. The slides were washed twice in PBS-triton for 10 min each. The third wash included bisbenzamide (Sigma-Aldrich, UK) (diluted according to the manufacturer's recommendation) which was incubated with the slides for 5 min. The slides were mounted with coverslips using ProLong Gold Antifade Mounting Medium (Invitrogen).

3.Alternative immunostaining method:

The slides were left to air dry for several hours. They were then washed with 1XPBS, 3 times at 5 min each Block with 5% Normal Donkey Serum in 0.3% Triton X-100 in PBS, incubated for 1 hr at room temperature. Primary antibody, diluting with 1% Normal Donkey Serum in 0.3% Triton X-100 in PBS was then prepared and these were washed briefly. The primary antibody was applied and left to incubate overnight at room temperature. They were then washed with 1XPBS, 3 times at 5 min each. The secondary antibody was then applied and incubated for 1 hr at room temperature. The slides were washed 1X PBS, 3 times for 5 min each. Apply DAPI, 1:5000 dilutions in PBS and incubate for 1 minute in the dark. Wash with 1X PBS, 3 times for 5 min each. They were again washed 1X TBS, 3 times at 5 min each, mounted with Vectashield, covered with a coverslip and sealed.

2.9 Immunocytochemistry (ICC)

Immunocytochemistry is a universal laboratory practice used to stain cells with specific antibodies raised to identify antigens. There are a wide variety of fixatives commonly used. It allows for identification of expression of antigens in particular cell types.

2.9.1 Methanol Fixation

Methanol precipitates proteins, which preserve cytoskeletal proteins well, but not the morphology of the cell. Ice cold methanol (-20˚C) was added to each well (0.5ml per well) and the plate was transferred immediately to -20˚C and incubated for 5 min. The methanol was discarded and the cells rehydrated with 1X PBS buffer for 5 min at room temperature. The cells were then blocked with 1X PBS – 0.5% BSA – 20mM glycine – 0.1% NaN_3 for at least 15 min before proceeding with staining, or stored at 4˚C.

2.9.2 PFA/Triton X-100 Fixation

Paraformaldehyde forms cross-links between proteins and thereby preserves the morphology of the cell. Cells were fixed for 20 min in 3% PFA in 1X PBS, at room temperature. The cells were permeabilised using 1X PBA – 0.3% Triton X-100 – 0.3% BSA, (300µl per well) and incubated for 5 min at room temperature. The cells were washed twice with 1X PBS – 0.5% BSA – 20mM glycine – 0.1% NaN_3, and incubated for at least 15 min before proceeding with staining or stored at 4˚C.

2.9.3 Immunofluorescent Staining

2.9.3.1 Cell Line-Y79

Media was vacuumed off and cells were washed 3 times in 1X PBS. The cells were fixed in 4% PFA in PBS for 15 min, and washed 3 times in PBS. Cells were permeabilised in 0.05% triton X 100 (Sigma, UK) for 10 min. Cells were then washed 3 times in PBS. 5% BSA (bovine serum albumin) in PBS was used as a blocking reagent, blocking for 30 min. Primary antibodies were diluted according to the manufacturer's instructions in blocking solution (1% BSA in PBS). Cover-slips were incubated with the primary antibody at room temperature for 1 hr. Cover-slips were washed 3 times in blocking solution before incubation with the corresponding secondary antibody (diluted 1:300 in blocking solution) for 45 min at room temperature in the dark. The cover-slips were washed 3 times in blocking solution and briefly stored in PBS before mounting onto slides using ProLong Gold Antifade Mounting Medium (Invitrogen).

2.9.3.2 General Growth

Cells were grown in T10 flasks or 6-well cell culture plates. Cells were grown in RPM1, supplemented with 20% heat inactivated stock fetal calf serum (First link UK Ltd) and 5ml penicillin/streptomycin antibiotic mixture. Media was changed approximately twice a week. Cells were passaged according to confluence/density. All work was carried out under sterile conditions, under a fume hood

2.9.3.4 Cell Passage

The media was aspirated and cells were washed once in sterile1X PBS. A mixture of trypsin/EDTA was used to detach the cells from flask; 3ml of pre-warmed trypsin/EDTA (37˚C) was added on top of the cells and left for 2-3 min at 37˚C to ensure that all the cells had detached. Fresh media (19ml, pre-warmed to 37˚C) was added to a fresh cell culture flask, and 10ml to the old (containing trypsinised cells) – the addition of fetal calf serum helping to quench the trypsin activity. Of the cells resuspended in media, up to 1ml was then transferred the new media and left at 37˚C. The old cell/media suspension was removed by suction.

2.9.3.5 Freezing Cells

Cells were pelleted by centrifugation at 800^{min-1} rpm for 5 min and resuspended in freezing medium (Gibco). The cells were then aliquoted into 1ml cryotubes and frozen slowly to avoid the formation of ice crystals; and placed at -80˚C. When frozen, they were moved to liquid nitrogen storage.

2.10 Zeiss LSM 510 Protocol

The Zeiss LSM 510 has 3 single photon lasers (Ar, HeNe 543, HeNe 633) which allow the user to select lines of 458, 488, 514, 543, 568, and 633 nm. The Zeiss LSM software provides a wide range of image processing functions, including all standard

2D/3D (stereo, projection) functions, digital processing of voxels and 3D measurement functions and parameters.

2.10.1 Turning on the lasers

ACQUIRE – LASER was selected this opens a laser control window with a list of available lasers. The laser needed to excite the dye of the labelled eye piece was turned on. When laser was set to "ready", the output intensity was turned at 75%.

2.10.2 Viewing specimen through eyepieces

2.10.2.1Send light to microscope

The ACQUIRE button was switched on in the toolbar of the main menu. The VIS button was then switched to on. The specimen was placed on the stage with the coverslip facing down and is held in place by two clamps.

2.10.3 View your specimen under transmitted light

The MICRO button was clicked and the Axiovert Control window appeared. The reflector turret position was turned to "none" and the transmitted light was switched "On". An objective was selected in the objective panel. The intensity switch was moved in front of the microscope until LED above the condenser read 2-3V. The specimen was found in XY. To find in Z focus, the large focusing and small focusing knob was used.

2.10.4 View the fluorescence signal of your specimen

The transmitted light in Axiovert Control was unchecked. Filter sets were selected for DAPI = Fset01. FITC = Fset09. Rho/Cy3 = Fset15. Once the appropriate field was selected, the Axiovert Control window was closed and the LSM button was changed for confocol scanning.

2.10.3 Scanning an Image

2.10.3.1 To collect an xy section (*Scan control)*

The acquire button was selected and scanning began. Under channel setting, various channels were selected based on the fluorophores used.

2.10.3.2 To acquire a Z-series stack

The Z-STACK button collects a series of XY-images along the Z axis. The Z-stack button was pressed to open the Z-settings window. The First/Last tab was selected. The CONT button was selected to start continuous scanning. While scanning, the manual focusing knobs were used to focus on one edge of the specimen. The MARK FIRST button was selected to set the limit of the stack. Focus in the opposite direction until the other end of the specimen was reached. The MARK LAST button was selected to set the end limit. The STOP button was used to cancel fast scans.

2.10.3.3 To collect a Time Series

Either a single or Z-series collection in Scan Control were set up. TIMESERIES was then selected, under which a series of time points were entered.

2.11 Electron Microcopy (EM)

The electron microscope is a high powered microscope using a beam of electrons to give a high magnification image of a specimen. This procedure was undertaken with the guidance and help of Dr. Peter Munro, Institute of Ophthalmology, UCL, London. Sections were cut and prepared by Dr. Munro. EM uses beams of highly energetic electrons to examine objects on a very fine scale.

2.11.1 Preparation of Tissue for EM sectioning

2.11.1.1 Fixation of tissue for immunogold labeling

Three fixatives are commonly used when fast (high pressure) freezing is not an option. These are 2-4% PFA, 0.1-1% glutaraldehyde (GLA) or a mixture of the two. Tissue was fixed in appropriate fixative – 1h minimum, 24h at 4ºC ideal with 4% PFA. Pieces were cut no larger than 3mm with a clean alcohol rinsed razor blade. Pieces were kept in storage solution until required.

2.11.1.2 Gelatin embedding

Tissue processing procedure involves the perfusion fixation with 4% paraformaldehyde in 0.08M sodium phosphate/cacodylate pH 7.4 for 2h at room temperature. Pieces were washed x3 in 0.08M sodium cacodylate + 50mM glycine. Pieces were embedded in 10% gelatine in Phosphate buffer + 5% sucrose and infiltrated in 2.3M sucrose overnight at 4°C.Mount and freeze.

2.11.1.3 Formvar coating of EM grids

Mesh hexagonal EM grids were cleaned by placing them in a 200-mL glass flask, rinsing them with 2% aqueous acetic acid. They were three times with dry acetone and dried the grids by placing the flask on a warm plate. A microscope slide was briefly immersed in dry ethanol and left to dry. The dry slide is dipped into a solution of 2% Formvar in chloroform and withdrawn carefully, to ensure a thick Formvar layer. The slide is left to dry by propping it vertically in a clean environment. A deep dish is filled with distilled water and left to stand below a diffuse light source so that the light reflects off the water surface. The water is cleaned by passing a lens tissue over the surface. A razor blade is used to score around the edge of the slide. The formvar-coated slide is used at a 45 angle at the edge of the glass dish and touched the bottom of the slide to the water surface. The glass slide was slowly and carefully put into the water while maintaining a 45 angle. The Formvar film started to detach from the glass surface and float onto the water surface. The glass slide was removed from the water when the Formvar film has completely detached from the glass surface. The floating film was examined for dirt particles. The cleaned EM grids were placed onto the floating film using a clean, sharp fine forceps. The grids were placed so that the shiny surface of the grid is facing up and the dull side is on the film. The glass slide were dried and used for casting the film and stick a 1-inch by 2.5 –inch white address label on its surface. The glass slide was held vertically over one edge of the floating film and pushed into the water so that the film floated on the water surface stuck to the address label on the glass side. Once the glass slide had been totally immersed, it was removed from the water making sure the grids remain attached and left to dry. The grids were coated with a thin

layer of carbon using a vacuum evaporator. The final carbon layer should be thin enough to form a light gray coating on a sheet of white paper.

2.11.1.4 Cryoprotection and Freezing

Cubes of tissue pellets were transferred to 1mL of 2.3 *M* sucrose in PBS in an eppendorf tube sealed and left overnight on a rotating mixing wheel at 4°C. The sucrose solution infiltrated through the specimen. Each infiltrated specimen block was transferred onto a clean metal specimen pin, excess sucrose was removed with a damp filter paper and the specimen was immersed in liquid nitrogen until freezing was complete.

2.11.1.5 Specimen transfer to precooled cryochamber

The cryochamber of the ultramicrotome was precooled to -80 °C. The diamond cryotrim tool was inserted and allowed the chamber to reach equilibration temperature. Any tools inserted into the cryochamber must be pre-cooled in liquid nitrogen. The specimen pin was placed into the specimen holder and locked in place using minimal force.

2.11.1.6 Trimming the Specimen Block

The antistatic line was turned on to full power. The front of the trim tool was advanced to the surface of the block until it was just touching the front of the block. The ultramicrotome was set to cut sections of 500 nm and the machine was left to section at the maximum speed setting. Once a substantial portion of the block had been smoothed away, sectioning stopped. This is known as “facing off”. The edge of the trim tool was

used to shape the sides of the block into a "mesa" protruding approx 1 mm wide and 2 mm long. Each side of the block was cut until a protruding mesa was formed with smooth sides.

2.11.1.7 Sectioning

The diamond knife was mounted in its holder, placed in the cryochamber and left to cool down and equilibrate at -120°C. The trimmed block was moved to the knife edge and cutting window set so that the section stroke started just above the knife and ended just below the knife. The antistatic line was set to 50% power or less. The ultramicrotome was set to cut 80-nm thick sections. The rough advance was used to move the knife close to the block. The knife continued moving to the block using the fine advance control until the two were almost touching. When it looked like the block and knife were almost touching, automatic sectioning began and sections to appear on the knife. The cutting speed was set to between 0.6 mm/s and 2 mm/s. The antistatic line was turned off. The sections were pulled from the edge of the knife with a clean eyelash probe and arranged in groups on the knife surface.

2.11.1.8 Section Retrieval

A 2-mm diameter loop was dipped in a mixture of one part 2% methyl cellulose: one part 2.3 M sucrose and lifted a drop out on the loop. The methyl cellulose/sucrose mixture was kept in a small tube on ice. The loop containing the sucrose/methyl cellulose drop was transferred to the cryochamber and maneuvered so the loop is just above a group of sections. The loop is moved down until the sections touch the drop of

liquid and immediately the loop is removed from the cryochamber. Sections adhered to the bottom surface of the drop of sucrose/methyl cellulose in the loop.

The liquid in the loop is thawed and placed, section side down, onto a Formvar/carbon coated EM grid at room temperature. The drop is then left to dry on the grid until it is ready for labeling.

2.11.1.9 Storing Sections

After the sections were placed onto the grids, they were stored for a short period of time by placing upside down onto 2% aqueous gelatin. The gelatin was first solidified by cooling to 4°C and the grids were placed, section side down onto a solid gelatin surface. To remove the grids for immunolabeling, the gelatin was warmed to 37°C for 30 min and taken off the grids with a wire loop. Long-term storage of sections can be achieved by leaving the grids attached to the glass slide.

2.11.2 Protein A gold procedure

Grids were transferred face down to 2% gelatin-PBS gels on ice – then warmed to 37°C. Grids were placed through 4x 3mins of PM+0.15M glycine (100ul) droplets. Grids were washed 2 x 5 mins with PBS+1% BSA. They were then incubated with primary antibody and diluted with PBS+1%BSA.Grids were washed 4 x 2mins with PB/0.15Mglycine. Grids were washed 5 x mins with PBS/0.1%BSA and incubated for 20 min on 10ul drops of Protein A diluted in PBS/BSA. Grids were washed 7 x 2 mins

with PBS+0.1% BSA and fixed for 5mins with PBS+1% glutaraldehyde. Grids were then washed 7 x 2mins with ultrapure water.

PROTOCOLS UNDERTAKEN BY COLLABORATORS

2.12 Comparative genome hybridization

Comparative genomic hybridization (CGH) is a comprehensive molecular-cytogenetic method for the analysis of copy number changes (gains/losses) in the DNA content of cells (Carter 2007). Genomic DNA from the tissue of investigation and normal DNA (reference tissue) are differentially labeled and simultaneously hybridised to normal metaphase chromosomes. The tested DNA is labeled with biotin and the reference DNA is labeled with fluorescein isothiocyanate. CGH detects only unbalanced chromosomal changes. However, structural aberrations such as balanced reciprocal translocations or inversions cannot be detected, as they do not change the copy number. CGH is also capable of detecting gain, loss and amplification of the copy number at the level of chromosomes. Detection of amplification is sensitive down to less than 1 Mb (Carter, 2007). One must take into consideration that while CGH is sensitive to amplification, it is an order of magnitude less sensitive to loss (Reference). The Sanger centre undertook this experiment using the Whole Genomic TilePath (WGTP) array.

2.12.1 Agilent Human Genome CGH Array

The Agilent Human Genome CGH Microarray Kit 244A is a high-resolution tool for genome-wide DNA variation profiling without amplification loss. Sufficient probe coverage spans both coding and noncoding regions. Probe design and selection have been carefully optimised and validated for maximal sensitivity and specificity. The enhanced detection resolution enabled through the high-density microarrays carry out the highest accuracy in mapping chromosome breakpoints and identification of micro-variations that are not detected by other arrays.

Protocols for WGTP and 224k are similar apart from a few minor adjustments:

2.12.2 WGTP array hybridization /244k array

Test and reference DNA samples were differentially labeled using the Bioprime labeling kit (Invitrogen Carlsberg, CA, USA) with modifications of the nucleotide mix. Briefly, a 260 µl reaction is set up containing 300 ng of DNA and 120µl for WGTP or 60 µl for 224k array of 2.5x random primer solution. After denaturing the DNA for 10 min at 100°C, 30 µl (WGTP) or 15 µl (244k) of 10x dNTP mix (1 mM dCTP, 2 mM dATP, 2 mM dGTP, and 2 mM dTTP in TE buffer), 3 µl (WGTP) or 1.5 µl (244k) of 1 mM Cy5-dCTP or Cy3-dCTP (NEN Life Science Products), and 6 µl of Klenow fragment were added on ice to a final reaction volume of 300µl (WGTP) or 150 µl (244k). The reaction is incubated at 37°C overnight and stopped by adding 30 µl of stop

buffer supplied in the kit. Unincorporated nucleotides are removed by use of Microcon YM-300 Filter Devices (Millipore Co.), according to the suppliers' instructions.

Hybridisations were carried out on a Tecan HSTM Hybridization Station (Tecan Group Ltd.) using 63x20 mm chambers. Cy3 and Cy5 labeled DNAs were combined, precipitated together with 270 μg of human Cot1 DNA (Roche Diagnostics Ltd., UK) and resuspended in 165 μl of hybridisation buffer (50% formamide, 5% dextran sulfate, 0.1% Tween 20, 2x SSC and 10 mM Tris/HCl, pH 7.4, 10 mM Cysteamine). Pre-hybridisation solution was prepared simultaneously by precipitating 100 μl of herring sperm DNA (10 mg/ml, Sigma Aldrich, UK) and resuspending in 165 μl of hybridisation buffer.

The prehybridisation and hybridisation solutions were then denatured for 10 min at 72°C. The prehybridisation solution was injected into the Tecan chamber following instructions displayed on the station. During prehybridisation (45 min at 37 ºC), the hybridisation solution was incubated at 37°C. Hybridisation was carried out for 21 hrs at 37ºC with medium agitation frequency. Slides were washed with PBS/Tween 20/2mM cysteamine (wash time 0.30 min, soak time 0.30 min, 15 Cycles at 37ºC), 0.1 x SSC (wash time 1.00 min, soak time 2.00 min, 5 Cycles at 54ºC), PBS/Tween 20/2mM cysteamine (wash time 0.30 min, soak time 0.30 min, 10Cycles at 23ºC) and HPLC water (wash time 0.30, soak time 0.00, 1 Cycle at 23ºC) before drying for 2.30 min using nitrogen gas. All experiments were performed in duplicate with DNA labelling color reversal (dye swap).

2.12.3 WGTP data analysis

Array images were acquired using an Agilent laser scanner (Agilent Technologies, UK). Fluorescence intensities and log2 ratio values were extracted using Bluefuse software (Bluegnome Ltd). Spots with low signal intensities ("amplitude"<100 in both channels) or inconsistent fluorescence patterns ("confidence" < 0.5 or "quality" = 0) were excluded before normalising all log2 ratio values by blocks (sub-arrays).

Fusion of dye-swap results and subsequent analyses were performed using custom Perl scripts. The median of all ratio values was calculated chromosome by chromosome for each individual hybridisation. Each ratio was then normalised by the corresponding chromosomal median. The ratios of each clone in the two dye swap hybridisations were then averaged if replicate ratios differed by less than 50% (i.e. less than a difference of 0.585 on the log2 scale).

The 68.2th percentile of the absolute values for all combined ratios was then calculated chromosome by chromosome as an estimation of the standard deviation (SDe). The 68.2th percentile of absolute dye swap rations for each chromosome is calculated as an estimated standard deviation. The mean of these give global rp68, which give an indication of quality (Carter, 2007). Clones reporting replicates different by more than eight times the SDe were excluded from further analysis.

Dye-swap experiments were accepted for CNV calling only if the following criteria were fulfilled: (1) Global SDe < 0.06; (2) Global clone exclusion at 23ºC) and HPLC water (wash time 0.30, soak time 0.00, 1 Cycle at 23ºC) before drying for 2.30 min

using nitrogen gas. All experiments were performed in duplicate with DNA labeling color reversal (dye swap).

2.13 MLPA reaction

This was undertaken in our collaborators laboratory in Seville, Spain. Genomic DNA was isolated from leucocytes using standard methods. MLPA Kit reagent EK1 was obtained from MRC Holland (Amsterdam, Netherlands), using a PTC-100 thermal cycler (MJ research, Waltham, Massachusetts, USA). 100 ng of DNA were diluted in 4 ul dH_2O and denatured at 98 ºC for 5 min before adding the MLPA buffer and the synthetic probe mixes. The reaction mixture was denatured for 1 min and incubated at 60 ºC overnight (16 hr). Ligation was performed at 54º C and incubated for 15 min. In a following step, the ligase enzyme was inactivated by heating up to 98º C for 5 min. PCR amplification of the ligation products was carried out in a 33 cycles program (30 s at 95ºC, 30 s at 60ºC, and 60 s at 72ºC). The products were separated on an ABI model 3730 DNA analyser (Applied Biosystems) using the GeneScan-500 LIZ size standards (Applied Biosystems) and analysed by using the GeneMarker version 1.6 (Softgenetics).

2.13.1.1 Probe design

Polymorphisms particularly across the ligation site were compared with the SNP database at NCBI and UCSC. The probes of 60 or fewer bases were synthesized to a scale of 0.05 μmol and purified by one step HPLC, whereas the ones spanning more

than 60 bases were manufactured to the scale of 0.02 μmol with the 3` half-probes including a 5` phosphate group for ligation purposes (Operon Biotechnologies GmbH (Cologne, Germany).

2.14 Bioinformatics Tools

This area of research was undertaken by Professor Chris Ponting and Doctor Leo Goodstadt of Oxford.

2.14.1 BLAST

BLAST (Basic Local Alignment Search Tool) is a set of similarity search programs designed to explore all of the available sequence databases regardless of whether the query is protein or DNA. Using BLAST at the website at http://blast.ncbi.nlm.nih.gov/Blast.cgi the RP25 genomic interval was used as a query to search protein and translated nucleic acid sequence databases for homologs.

2.14.2 ExPASY

ExPASY found at http://www.expasy.ch/tools/ was another tool used to further analyse the protein sequence of RP25. ExPASY is a database server containing access to many bioinformatics databases and molecular biology analytical tools. Such tools in the ExPASY database include

SWISS-PROT- protein sequence database that gives annotations such as protein domain structure and post translational modifications.

PROSITE- this also gives details on various protein domains and families. It has detailed profiles which help to identify to which known protein family a new sequence belongs to. Both of the above bioinformatics tools were used in this study to reveal the protein structure of the *RP25* gene. The Molecular Biology Server (BCM) Search Launcher was also used in this instance. This is an integrated set of World Wide Web (WWW) s that organizes molecular biology-related search and analysis services available on the WWW by function, and provides a single point of entry for related searches.

BLAT (Blast-like alignment tool) is a sequence alignment server and is used for sequence comparisons against an entire genome. BLAT is found at http://genome.ucsc.edu. In this case it incorporated the genomes of species such as rhesus monkey, dog, cat and mouse.

2.14.3 Conserved Domains Tools

The domain architecture of what was predicted using tools including SMART and hmmsearch.

2.14.3.1 SMART protein domain database server

SMART (a Simple Modular Architecture Research Tool) (http://smart.embl-heidelberg.de/) allows the identification and annotation of genetic domains and the analysis of domain structure. More than 400 domain families found in signaling, extracellular and chromatin-associated proteins are detectable. These domains are extensively annotated with respect to phyletic distributions, functional class, tertiary structures and functionally important residues (Schultz *et al.,* 2000).

2.14.3.2 HMMsearch

HMMsearch is a suite of programs that uses Hidden Markov Models (HMMs) to describe the profile of a multiple sequence alignment. HMM is a finite set of states, each of which is associated with a probability distribution. In a particular state an outcome or observation can be generated, according to the associated probability distribution. It is only the outcome, not the state visible to an external observer and therefore states are ``hidden" to the outside; hence the name Hidden Markov Model. HMMsearch profiles can in turn be used to perform incredibly sensitive database searches and to merge evermore distantly related sequences into the original alignment. One can also search a protein sequence against a profile library to detect known domains.

2.14.4 Orthology and phylogenetic tree

Homologs of the RP25 protein were identified using BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on). This

simply compares the protein sequence to the protein database available. The domain sequences were manually aligned and the protein distances were calculated.

The protein distances were calculated using prodist from PHYLIP. Prodist uses protein sequences to compute a distance matrix, under three different models of amino acid replacement. The distance for each pair of species estimates the total branch length between the two species. The distance matrix programs used in this case were neighbour and consensus.

The first step to constructing a phylogenetic tree is to calculate distance matrices for the multiple sequence alignment. The neighbour-joining tree from the distance matrix is calculated. Inspect the tree by viewing with an outtree file, once tree is satisfactory, calculate the bootstrap values. Bootstrap values are a computer-based technique for assessing the accuracy of almost any statistical estimate (Efron *et al.,* 1996).

Summary of Materials

Item(s)	Company
1X TAE	Invitrogen
10X Loading Buffer, Bromophenol Blue	Sigma-Aldrich
10X NH_4 PCR buffer	Bioline
10X TRIS + glycine +SDS	National Diagnostics
Acrylamide	National Diagnostics
Agarose	Sigma-Aldrich UK
Ampicillin	Sigma-Aldrich UK
BigDye® Terminator v3.1 Cycle Sequencing Kit	Applied Biosystems
Bovine Serum Albumin, BSA	Sigma-Aldrich UK
Bromophenol blue	Sigma-Aldrich UK
CACO2 Cells	Personal communication
Calf Intestinal Phosphatase	Roche Diagnostics GmbH
cDNA	Clontech, UK; made in-house
Chloroform	BDH-Analar UK
DEPC	Difco Laboratories
Demethyl Sulfoxide, DMSO	Sigma-Aldrich UK
D-MEM Liquid Medium	Gibco BRL
D-MEM/F12	Gibco BRL
Dnase	Ambion
dNTPs	Promega UK
EDTA	Affymetrix
Ethanol	BDH-Analar UK
Ethidium Bromide	Sigma-Aldrich UK
Fetal Calf Serum	First Link UK Ltd
Formvar	AppliedChem
Freezing medium	Gibco BRL

Gelatine	Sigma-Aldrich UK
Glutaraldehyde	Sigma-Aldrich UK
Glycerol	Sigma-Aldrich UK
Glycine	ICN Biomedicals, Inc
Glycogen	
Hydrochloric acid	
Isopropanol	BDH-Analar
Kanamycin	Sigma-Aldrich UK
KOD Hot Start DNA Polymerase Kit	Toyobo Novagen
Methanol	BDH-Analar
Methly cellulose	Sigma Aldrich UK
Magnesium Chloride, $MgCl_2$	Bioline
Normal Goat Serum, NGS	Sigma-Aldrich UK
Paraformaldehyde	Sigma-Aldrich UK
Penicillin/Streptomycin	Gibco BRL
Phenol	Sigma-Aldrich UK
Phosphate Buffered Saline tablets	Oxoid UK
Precision Plus Protein dual color standards	Bio-Rad Laboratories
Primers	Sigma-Genosys
Prolong Gold anti-fade reagent for mounting	Invitrogen
QIAquick PCR Purification Kit	QIAgen
Restriction Enzymes & buffers	Promega, New England Biolabs
RPMI-1640 Medium	Gibco BRL
RNasin	Ambion
Sodium azide	Sigma-Aldrich UK
Sodium cacodylate	Sigma-Aldrich UK
Sodium dodecylsulfate, SDS	Sigma-Aldrich UK
Secondary Antibodies	Jackson Immunoresearch Labs Ltd USA

SmartLadder - Molecular Weight Marker	Eurogentec
S.O.C. Medium Liquid	Invitrogen
Sodium Chloride, NaCl	Fischer Scientific
Sodium phosphate	Sigma-Aldrich UK
Sucrose	Sigma-Aldrich UK
Taq polymerase	Bioline
Trizma Base	Sigma-Aldrich UK
Triton X-100	Sigma-Aldrich UK
Trizol	Invitrogen, UK
Trypsin/EDTA	Gibco BRL
Tryptone	Merck
Tween-20	Sigma-Aldrich UK
Vectashield mounting medium with DAPI	Vector Laboratories

CHAPTER 3

Candidate gene exclusion and identification of the *RP25* gene

3.1 Background of screening candidate genes in the RP25 interval

Prior to commencing my PhD studies, 37 of the 110 genes in the RP25 interval (34 Mb) were screened by my colleague in London and by our collaborating group in Seville, Spain. This, together with previous reports excluded 54 genes within the RP25 interval. All screened genes were thus excluded as disease causing for RP25 leaving 56 genes as possible candidates. Details on the RP25 locus and linked Spanish families can be found in section 1.12. Selection of candidate genes to screen was based on known gene function, tissue expression pattern and/or the genetic data published on the genes. Two main approaches were used to prioritise candidate genes for mutation screening; which are detailed below.

3.1.1 Functional candidate gene approach

The initial essential stage in identifying candidate genes was the selection process. Genes in the interval may already be implicated in a role related to retinitis pigmentosa or may be highly expressed in retinal tissue. For this approach, an interactive database

called EysSAGE (Serial Analysis of Gene Expression) was used to identify tissue expression of various genes within the RP25 interval. This database queries human retina and RPE gene expression. Some of the 54 genes that were selected as good candidate genes using this approach are as follows: *GABRR1* and *GABRR2*, which are both expressed in the retina and encode rho1 and rho2 subunits of the C type receptor for the ã-aminobutyric acid (GABAc receptor) (Marcos *et al.,* 2000). *ELOVL4* is another gene selected based on its composition and expression in photoreceptor cells (Li *et al.,* 2001) (Leonard *et al*., 2000); which has also been previously implicated as the causative gene for STGD3 and ADMD (Zhang *et al.,* 2001). *SMAP1* (stromal membrane associated protein 1) is another example of a gene that is located within the RP25 locus and is expressed in the retina. Similarly, *GLULD1*, glutamate-ammonia ligase (glutamine synthase) domain containing 1, plays a key role in the uptake and metabolism of glutamate in the retina (Barragan *et al.,* 2005). *RIM1*, encoding a presynaptic protein involved in the glutamate neurotransmission, is the gene responsible for autosomal dominant cone-rod dystrophy CORD7, whose locus overlapped partially with the RP25 locus (Barragan *et al.,* 2005).

3.1.2 Positional candidate gene

A positional candidate gene list was produced based on the genetic data obtained from the Spanish families and other ethnic families linked to the interval (section 1.12). The interval flanked by markers D6S282 to D6S1976 was divided into four regions (Figure

3.1). All four regions were thought of as being important and so to devise a strategy plan, a thorough analysis of known functions and expression patterns of the genes had to be explored. This involved performing a functional candidate gene approach. The four regions analysed were as follows:

(1) Region A is where the Spanish families and the Pakistani family overlap and is between microsatellite markers D6S257 and D6S1053, spanning approximately 8 Mb and has 15 genes mapped to the region.

(2) Region B is where six Spanish families are genetically linked to the region between microsatellite markers D6S1053 and D6S1557 spanning approximately 6 Mb and has 11 genes.

(3) Region C is where all Spanish families map, between microsatellite markers D6S1557 and D6S421 spans approximately 1 Mb and has 7 genes.

(4) Finally, region D is where the Spanish families overlap with the LCA5 (Leber congenetial amarousis 5) locus, between microsatellite markers D6S421 and D6S1644 spans ~ 18 Mb and contains 79 genes.

Up to the point of project commencement, no gene had been identified for LCA5, which is also in the same chromosomal region (section 1.12). This led to the possibility that the same gene could be causative for both RP25 and LCA5. This has previously been reported with genes such as *CRB1* (Richard *et al.,* 2006), *ABCA4* (Allikmets *et al.,* 1997) and *RPE65* (Redmond *et al.,* 1998) which were reported to cause both LCA and RP phenotypes. In total, 54 genes had been screened in all four regions, including all of the genes referred to in section 3.1.1. Fourteen genes in region A had been selected as

possible candidates and then screened for mutations. In region B, five genes were also selected as possible candidates and further discounted once they were screened, leaving seven uncharacterised gene transcripts to be screened in the region. Seven genes were selected as good candidate genes and subsequently discounted from region C once all exons in these genes were sequenced. In region D, the largest region, 28 genes were selected as possible RP25 causing genes and disregarded as the causative RP25 gene after full exonic sequencing.

3.2 Sequencing and screening of genes

Initial screening began with the completion of genes which contained exons that were difficult to sequence. Numerous genes including those described in table 3.1 had exons which proved difficult to amplify. Many new primers for these exons were designed, as well as using alternative temperatures and magnesium concentrations during PCR amplification. It is extremely important that every exon is completely screened in order for the gene to be excluded as a possible disease gene. Once these genes were completely screened, the next stage of the project was to select candidate genes for sequencing.

Gene/gene description	**Gene size (Kb)/Position and region location**	**No.of exons**	**UniGene**	**Published information on gene**
ZNF451-Zinc finger protein 451	80.27/57.06-57.14/Region A	14	HS.485628	Co activator for steroid receptors.
COL19A1/NM_001858/Collagen, type XIX, alpha 1	345.71/70.63-70.97/Region B	51	HS. 444842	COL19A1 and COL9A1 were duplicated from the same ancestral gene of the FACIT family (Khaleduzzman et al., 1997).
COL9A1/Collagen, type IX, alpha 1.	87.03/70.98-71.06.	38	HS.590892.	A mutation in the COL9A1 gene was reported in a family with multiple epiphyseal dysplasia (Czarny-Ratajczak et al., 2001). A homozygous mutation was also reported in a recessive form of Stickler syndrome (Van Camp et al., 2006).
SH3BGRL2- SH3 domain binding glutamic acid-rich protein like 2.	72.37/80.39-80.47/Region C	4	HS.232772.	Reported as a novel human homolog of the SH3 binding glutamic acid-rich establishing a new family of highly conserved proteins related to Thioredoxin Superfamily (Mazzocco et al., 2002).
TTK-TTK protein kinase Dual specificity protein kinase	37.88/80.77-80.80/Region C	22	HS.169840	Has dual specificity protein kinase (tyrosine and serine/threonine residues) (Lindberg et al., 1993) Is required for centrosome duplication and for the normal progression of mitosis (Fisk et al., 2003).

3.3 Screening genes in region D

Once incomplete genes were completely screened, genes in region D of the RP25 locus were analysed. Seventy nine genes were identified between microsatellite markers D6S421 and D6S1644, region D (Figure 3.2) with 28 genes being previously screened and excluded as potential disease causing genes for RP25. In total, six candidate genes were selected for mutation screening (Section 3.2.1) (Table 3.2). These candidate genes were selected at random as all good candidate genes had been completely screened previously. Primer pairs, (Sigma Aldrich, UK) were designed for all 99 exons of the six genes (see appendix for primer design). Several novel SNPs were identified, along with previously annotated SNPs in these candidate genes (Section 3.2.2).

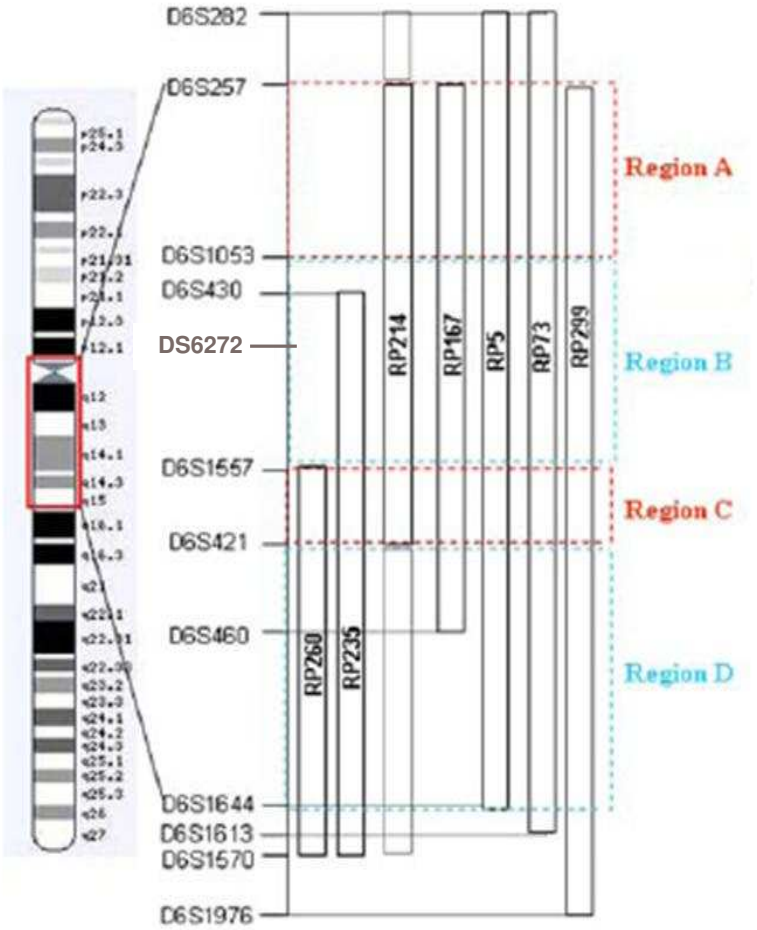


Figure 3.1 The *RP25* interval on chromosome 6. An overview of where the seven Spanish families lie in the interval and the corresponding microsatellite markers. The *RP25* interval was divided into four separate regions A-D. The genes been screened in this instance lay in region D between microsatellite markers D6S421 and D6S1644. Adapted from Abd El-Aziz *et al.,* 2008.

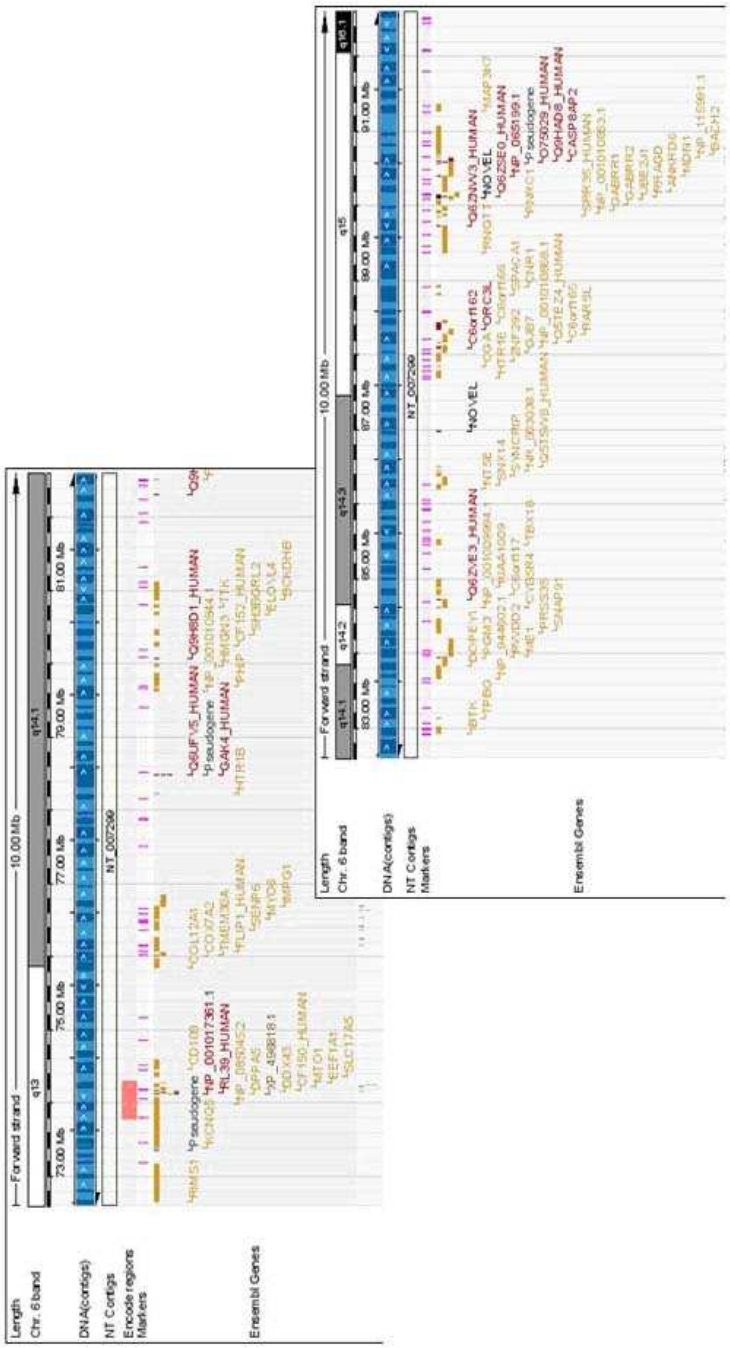


Figure 3.2 An Ensembl overview of the predicted genes in the RP25 D region. At the time of screening, 79 genes were annotated in the region.

Table 3.2 Genes screened for mutations

Gene Description	No. of exons	Position	Expression data
DPPA5 (Developmental pluripotency associated 5)	3	74.11-74.12	None
MTO1 (Mitochondrial translation optimisation 1 homolog)	13	74.22-74.26	Ubiquitously expressed
SENP6 (SUMO1/sentrin specific peptidase 6)	24	76.36-76.48	Ubiquitously expressed
ME1 (Malic enzyme 1)	14	83.97-84.19	Ubiquitously expressed
CYB5R4 (Cytochrome b5 reductase 4)	16	84.62-84.72	Ubiquitously expressed
SNX14 (Sorting nexin 14)	29	86.27-86.36	Ubiquitously expressed

3.3.1 Information on genes screened

DPPA5 (developmental pluripotency associated 5) is also known as ESG1 (PH34/ECAT2). This gene has been reported to have an important role in human embryonic stem and germ cells and also can be used as a marker of pluripotent stem cells (Kim *et al*., 2005). *MTO1* (Mitochondrial translation optimization 1) has been proposed as a candidate gene to modulate the effects of nonsyndromic sensorineural deafness mutation A1555G (Bykhovskaya *et al*., 2004). It is also suggested to be involved in the process of mitochondrial RNA modification, which is an important regulatory pathway for nonsyndromic sensorineural deafness. *SENP6* encodes an ubiquitin-like molecule (UBL), SUMO-1 UBLs are small proteins that are covalently conjugated to target proteins with important functional consequences such as protein degradation and/or change in localisation (Matic *et al*., 2008). *SENP6* (*SUMO1*) mainly modifies nuclear proteins, and participates in the modulation of ER-mediated corticotrophin-releasing hormone (CRH). This may be important for the regulation of the stress response (Zhu *et al*., 2008). It is reported to be highly expressed in organs, such as testis, ovary and prostate suggesting it may play a role in reproduction (Kim *et al*., 2000). *ME1* encodes a cytosolic NADP+-dependent enzyme involved in the regeneration of pyruvate from malate back to the mitochondria, forming a link between the glycolytic pathway and the citric acid cycle. By assisting the release of acetyl–coenzyme A (CoA) and NADPH from the mitochondria into the cytosol, it makes these compounds available for de novo fatty acid biosynthesis and other metabolic processes. *ME1* is considered lipogenic. Recently, it was identified as a primary candidate gene underlying a porcine quantitive trait loci (QTL) associated with back fat thickness

(Yang *et al*., 2009). *SNX14* is a G-protein signaling intracellular trafficking gene and is mainly expressed in neurons (Carroll *et al.,* 2001). *CYB5R4* (*NCB5OR*) is a flavohemoprotein that contains functional domains found in both cytochromoe b5 and CYB5 reductase. It is a single 58 kDa polypeptide composed of two well characterized domains tethered by a unique 90-residue hinge (Larade *et al.,* 2008).

3.3.2 SNPs identified in screened genes

In *DPPA5*, one novel SNP was identified in an affected member of the RP5 family c. 297-30 G>A. Since this synonymous SNP did not lead to an amino acid change, there was no requirement to check the allelic frequency (Figure 3.3). Several already annotated SNPs were also identified in members of all linked families. Three novel SNPs were identified in *MTO1*, along with several members having numerous already known SNPs. Both affected and unaffected members of family RP214 and the affected member of RP5 and RP73 all had individual novel SNPs (Figure 3.4). These were found at positions a) c. 879–46 A>G, b) c. 1201–59 C>G and c) c. 1409 + 14 T>A. All three novel SNPs were identified in the intronic regions therefore do not lead to any amino acid changes. However, these changes may be involved in modifying the normal splicing of the relevant candidate gene or by affecting expression levels through changes to putative regulatory elements.

Three novel SNPs were identified in *SENP6* in individuals of several families (Figure 3.5). The SNPs were as follows c. 1129 A>G – now a known SNP (rs16886792), c. 420 + 453G>T, c. 637–135 G>C. The first SNP was found in the affected member of RP73. The second SNP was identified in unaffected and affected

members of RP299. The final SNP was identified in both affected members of RP5 and RP73 respectively. Again all three SNPs were not pathogenic. Several SNPs, already identified, were found in all members screened. One Novel SNP in ME1; c. 302 + 21A>C (Figure 3.6) was identified in all members of RP5 which is now an annotated SNP (RS4706990) found on Ensembl. No novel SNPs or mutations were identified in either *SNX14* or *CYB5R4,* only several known SNPs. Therefore all of these genes were ruled out as possible candidate genes for the *RP25* locus.

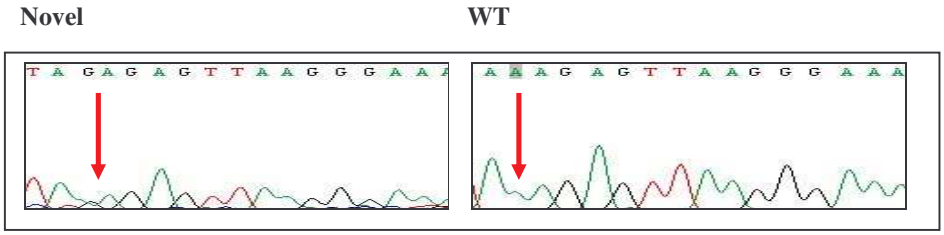


Figure 3.3: Electropherograms of the novel SNP c.297-30 G>A identified in affected member of RP5 in the gene *DPPA5* (See Automated Sequencing Section 2.4)

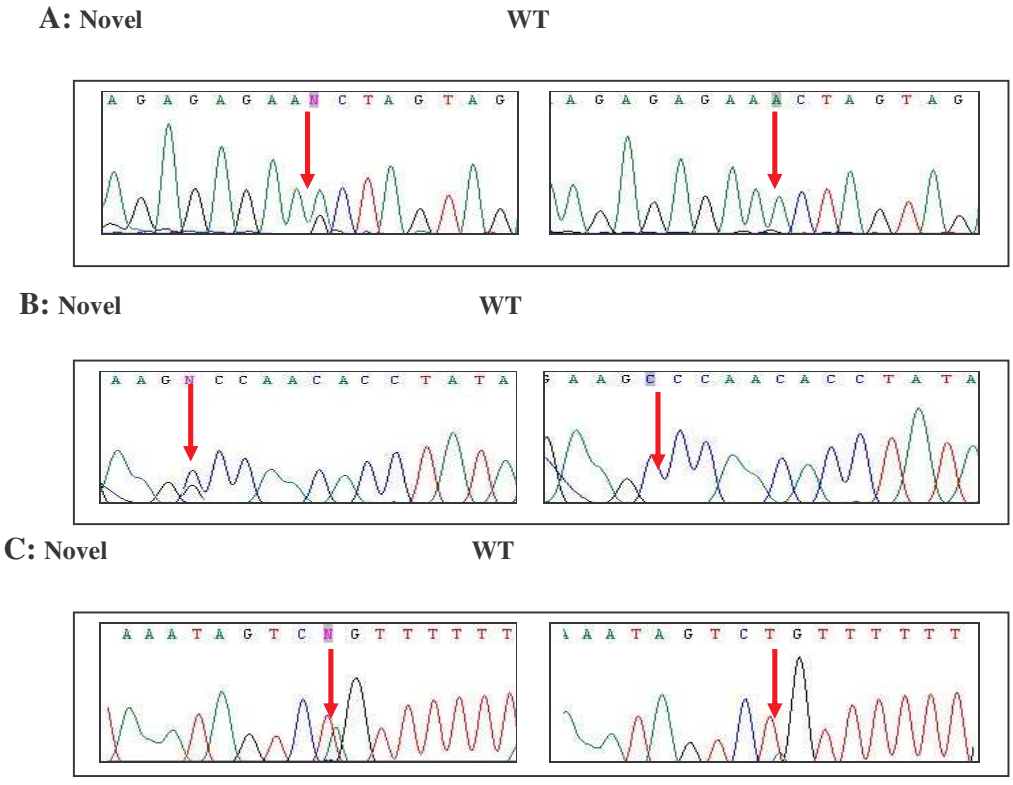


Figure 3.4: Electropherograms of SNPs identified in *MTO1* **A)** Novel SNP c.879-46 A>G identified in unaffected members of RP5 **B)** Novel SNP c.1201-59 C>G. **C)** Novel SNP c.1409 +14 T>A identified in affected and unaffected members of RP5 and RP73.

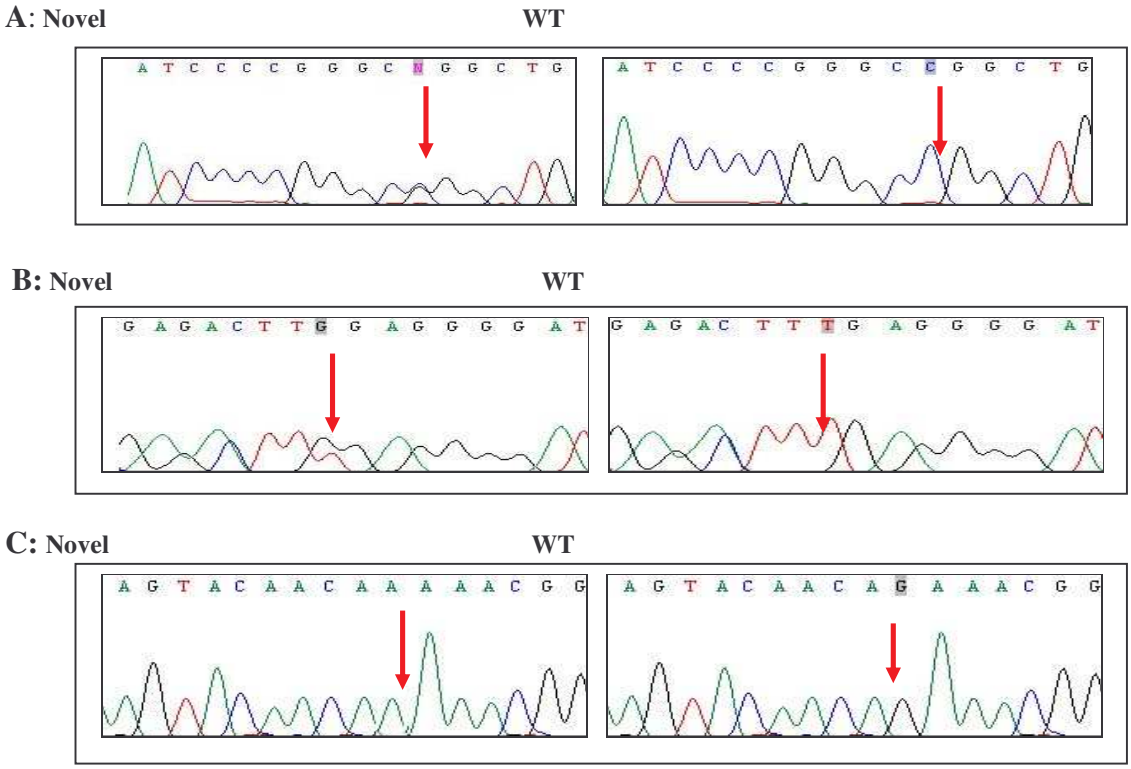


Figure 3.5: Electropherograms of SNPs in *SENP6*. **A)** Novel SNP c. 637-135 G>C identified in several families, including RP5, RP299 and RP167.**B)** Novel SNP c.420 +453 G>T identified on all members of RP5. **C)** SNP c.1129 A>G now a known SNP (rs16886792)

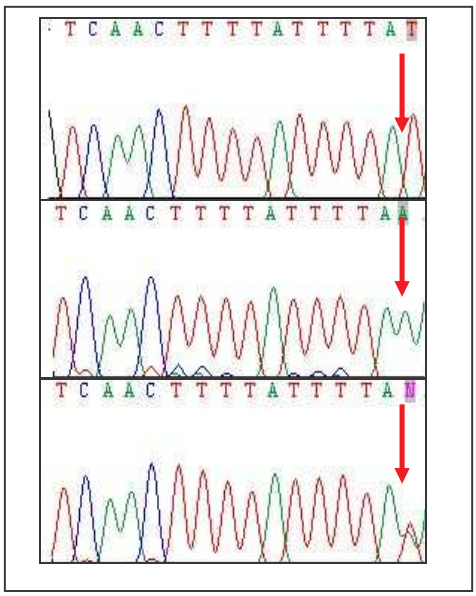


Figure 3.6: Electropherogram of SNP RS4706990 identified in *ME1* in RP167, RP214, and RP299.

3.4 Identification of the RP25 gene

3.4.1 Copy Number Variation (CNV) and Comparative Genome Hybridisation (CGH) arrays

In the previously analysed genes in the RP25 interval, all identified SNPs segregated with the genetic data of each of the families. However, one gene, *PRIM2A* came to attention when heterozygous SNPs were observed within a stretch of homozygosity in the affected members of two consanguineous families (RP5 and RP214). This did not correlate with the linkage data of these families and so it was postulated that a second copy of this gene existed in the same interval. This is also known as copy number variation (CNV), which would mean that amplification of two copies of the gene occurred at the same time. In addition to this, it would also be possible that in one of the copies of *PRIM2A* at a specified position, a sequence of dinucleotide AA would exist, and in the other copy a sequence of TT would exist. Through amplification of both copies, the two strands of DNA would lead to the existence of the identified heterozygous SNPs in the consanguineous families. These observations led us to undertake the CNV study.

In collaboration with Sanger Centre, Cambridge, six samples belonging to the RP5 family were subjected to a whole genome tilepath array analysis. This molecular-cytogenetic technique used for the analysis of copy number changes in DNA, detects only unbalanced chromosomal changes and identify abnormal regions in the affected genome. The results from this study revealed that a clone within the RP25 locus was deleted in all affected members of this family. A whole genome tilepath array was

undertaken on each of the members and as a result a deleted clone was identified in the affected members (Figure 3.7 and 3.8). The deleted clone was also identified in the carriers of this family but at a higher ratio than that off the affected members demonstrating that this mutation is present as a single copy, all unaffected members read with normal ratios for this deletion. A clone is called as having a copy number difference in the test compared to the reference. If the ration is >6x rp68 (Section 2.13) for an isolated clone or > 4x rp68 for a consecutive clone.

3.4.2 The deleted clone

The deleted clone, chr6tp-19C7 is located at 65.7 Mb within the 6q12 interval in all affected members of family RP5. A 95 Kb region of sequence proximal to Chr6tp-19C7 was not covered by the array used in this experiment. Hence, it was not possible to delineate the proximal breakpoints of the deletion since there is a gap in the clone coverage. The proximal clone Chr6tp-10D10 reports with a normal ratio, suggesting it is not affected by the deletion. The deleted clone does however overlap with a distal clone Chr6tp-10G7. The overlap with this clone spans 40 Kb, and also reports a normal ratio (Figure 3.9). This deletion is novel as it had not been previously reported as a known CNV and is a large deletion of 100 Kb in size. This deleted clone could pinpoint the gene responsible for RP25 as it segregates with the disease phenotype in the RP5 family.

A 244k Agilent array (Figure 3.10) and a Nimblegen data array were performed on the affected members to verify the segregation of this copy number variation in the RP5 family (Figure 3.11).The breakpoints for this deletion were identified later (Section 4.10) in the study using both MLPA (multiplex ligation probe dependent assay) and a 244k Agilent array. The most interesting aspect of the identification of this deleted clone was the precise mapping within the RP25 interval. The corresponding genes mapped in this region were a cluster of uncharacterised gene transcripts in region B of the RP25 interval (Figure 3.9).

3.5 Mutation screening of genes in region B

In parallel to CGH analysis, PCR amplification and sequencing of candidate genes in the RP25 locus continued. In the meantime, our Spanish collaborators had collected five more unrelated Spanish families who were linked to the RP25 locus (Barragan *et al.,* 2008). In addition to the originally linked families (RP5, RP73, RP167, RP214 and RP299), these five new families (RP328, RP349, RP355, RP377 and RP351) were incorporated into the study (Figure 3.12), with three of the ten families being consanguineous (RP5, RP167 and RP377). This further confirmed the high prevalence of RP25 in the Spanish population. The linkage of these new families also refined the RP25 interval from 16.1 cM down to 2.67 cM.

Based on the CGH array results and the mapping of five more Spanish families to the RP25 interval, sequencing of candidate genes moved from region D to B. Genes which

had been screened and previously excluded in region B are *COL19AL*, *EGFL11*, *BAI3* and *LMPRD1* (Table 3.3). Seven uncharacterised gene transcripts were left to be analysed, they are as follows *Q5T669*, *Q5T1H1, Q9H557_human, Q5TEL3_human, Q5TEL4_human, Q5VVG4_human,* and *Q5T3C8*. Six of these transcripts share a common epidermal growth factor (EGF-like) domain; however none were identified as having full length transcripts or had defined translation start or stop codons. All seven transcripts mapped in very close approximation to one another in the region of interest (Figure 3.13).

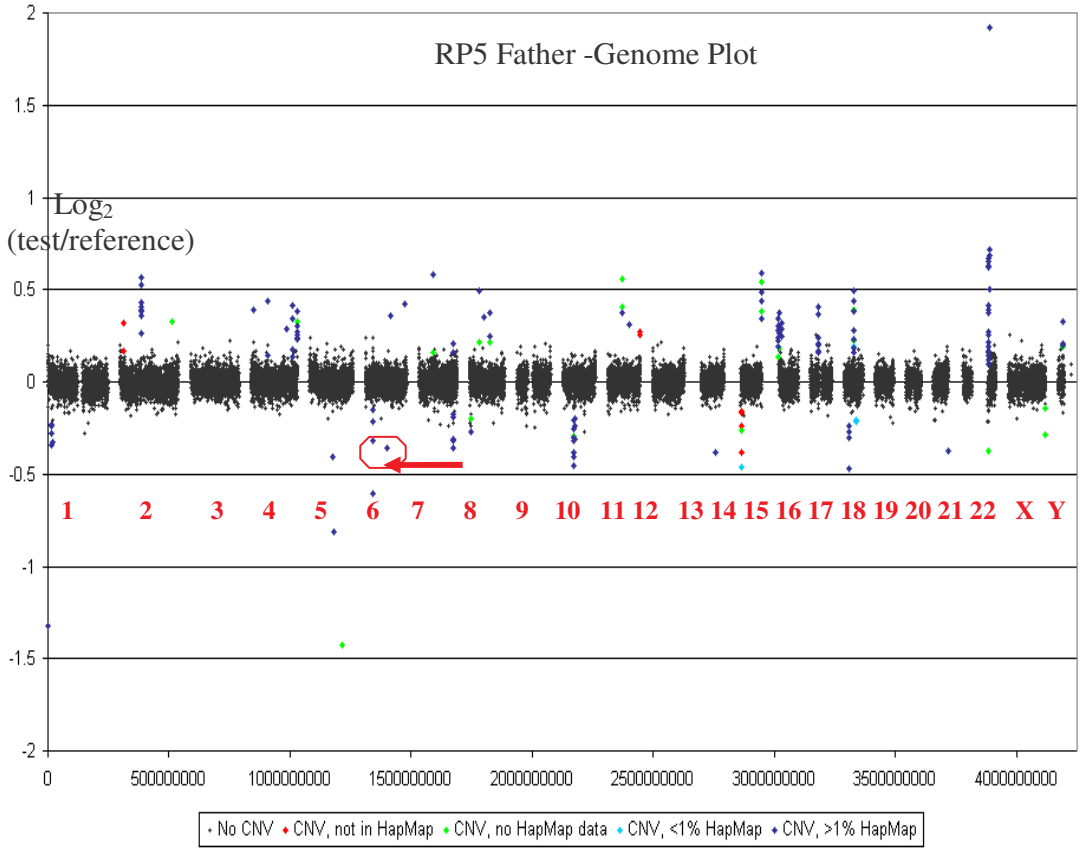


Figure 3.7 Whole Genome Tilepath Array data run on the father of the RP5 family. Each profile represents a single chromosome. The points in red highlight significant copy number variation (CNV), light blue represents regions of CNV variants in less than 1% of the population, blue represents regions of CNV variants in more than 1% of the population, and the points in green have no hapmap information. A deleted clone identified in father (carrier) has a ratio of -0.201 in chromosome 6 as depicted by the red arrow. Experiment was undertaken by Sanger Centre, Cambridge.

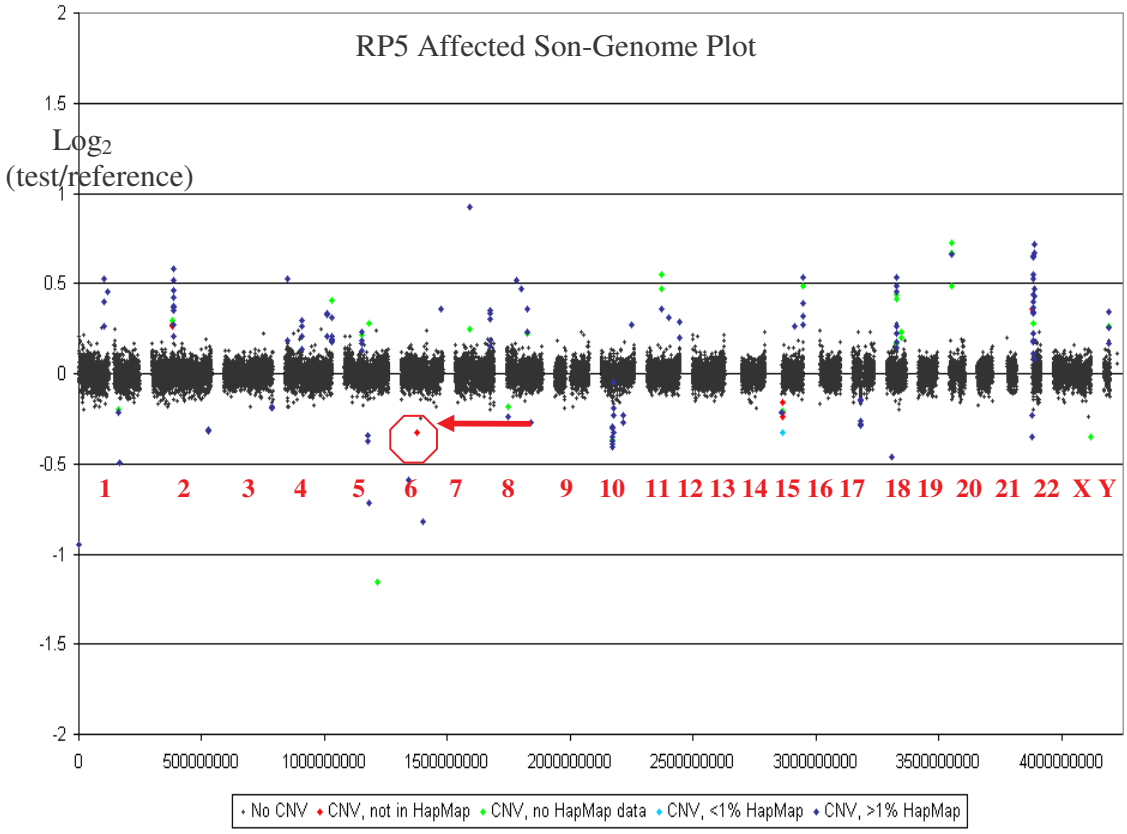


Figure 3.8 Whole genome tilepath array run on one affected son from family RP5, attention was on chromosome 6 where the RP25 locus maps and where a variation in CNV was identified. A copy number plot (red) is identified in chromosome 6 of this affected member of RP5; this deleted clone in this affected member has a ratio of -0.307.

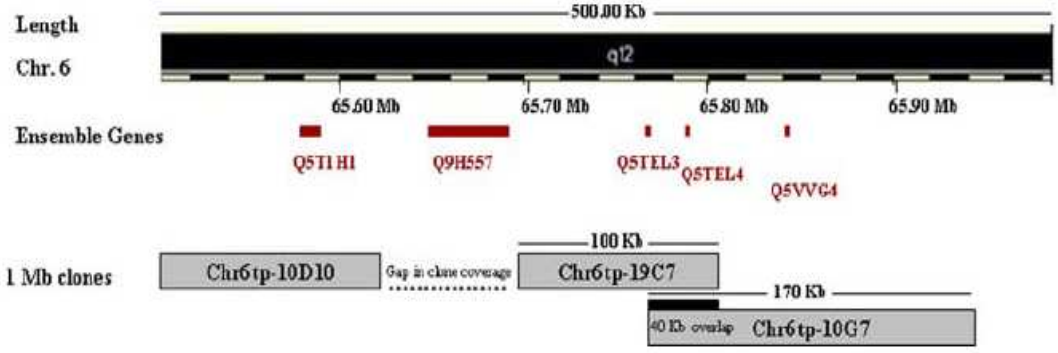


Figure 3.9 Identification of a deleted clone, Chr6tp-19C7 in an affected member of family RP5. This clone spans over 100Kb. A 95 Kb region of sequence proximal to Chr6tp-19C7 was not covered by the array used in this experiment. The nearest clone Chr6tp-10D10 reports with a normal ratio, suggesting it is not affected by the deletion. The deleted clone does overlap; however, with a distal clone Chr6tp-10G7. The overlap spans 40 Kb, and also reports a normal ratio (Adapted from Abd-El Aziz *et al.,* 2008)

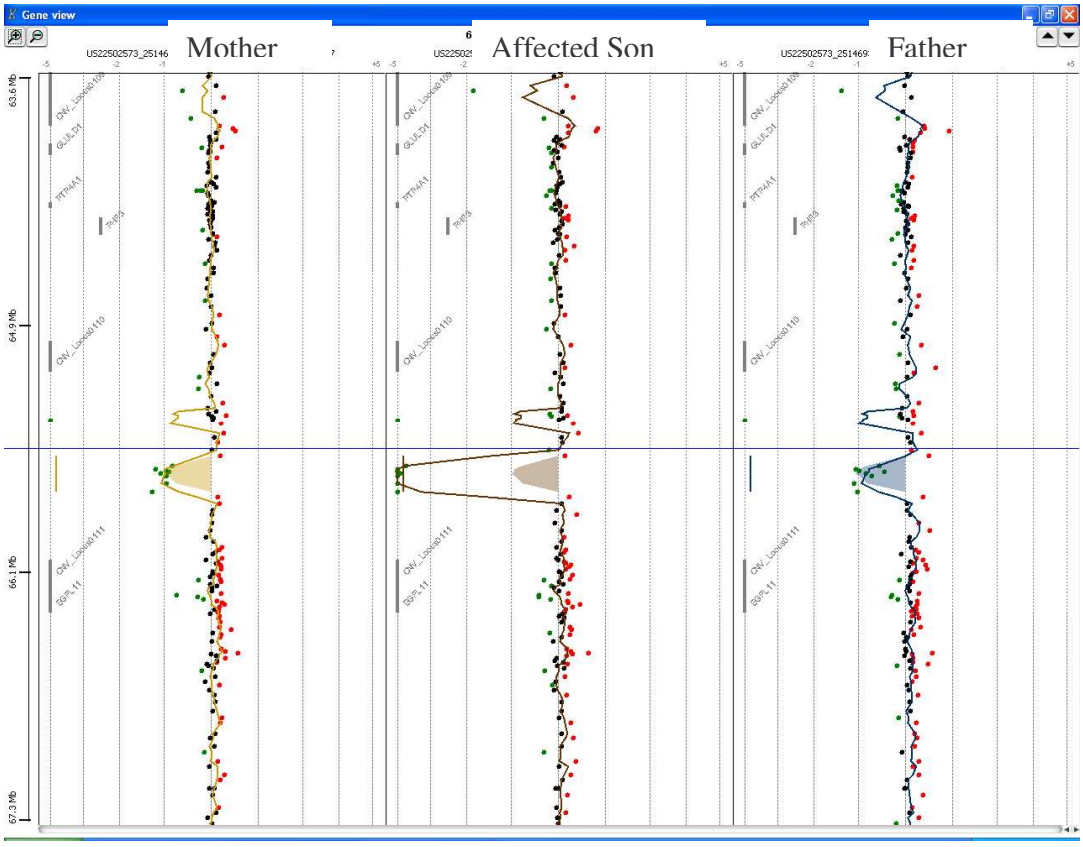


Figure 3.10 An Agilent array demonstrating the region where the deletion lies in the RP5 family. The height of the peak in the affected son of this family highlights the deletion of a clone in this interval. The heights of the peaks in both parents highlights that both parents have one allele with the same deletion. Experiment was undertaken by Sanger Centre, Cambridge.

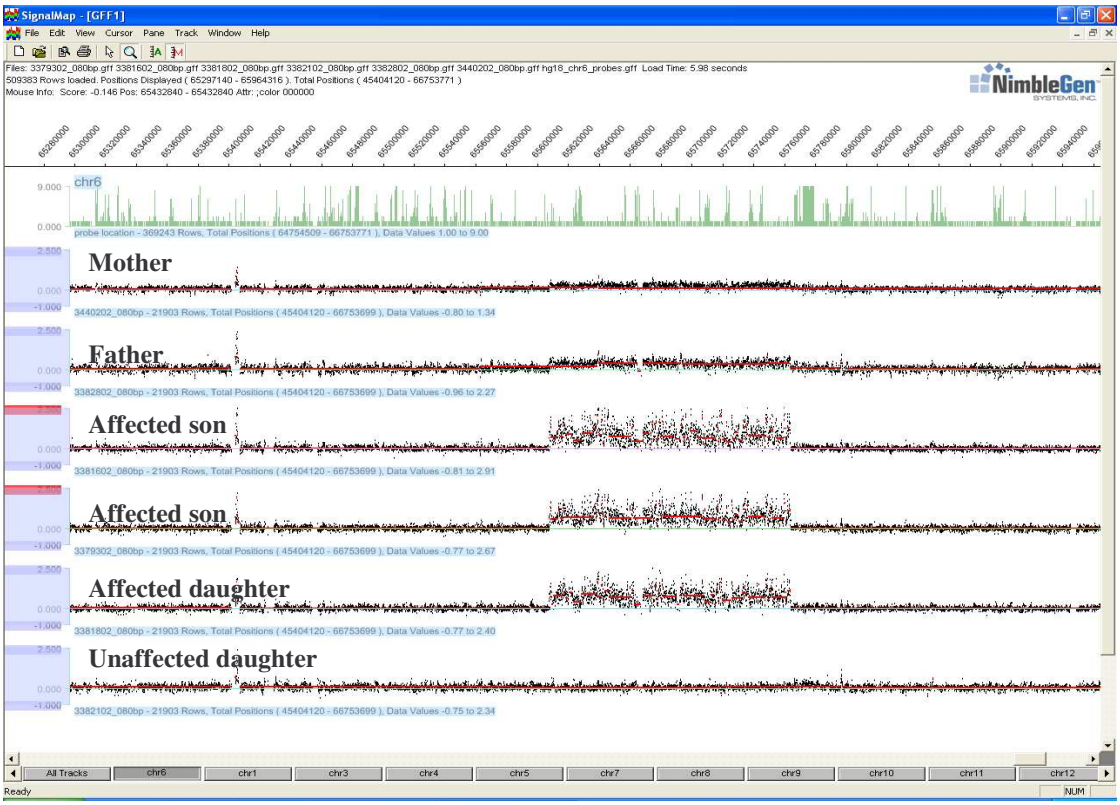


Figure 3.11 Depiction of Nimblegen data array run of RP5 family. The array shows the nimblegen array run on chromosome 6. This was designed to cover a 2 Mb interval from 64753791 and 66753791 Mb. The deletion is identified as a noisy signal. The Mother and Father both exhibit a run of chaotic peaks from approximately 6560000 to 6578000 Mb; both parents are carriers of this deletion. Both affected sons and affected daughter have a run of large chaotic peaks, as the deletion appears as a duplicate. Finally, the unaffected daughter of the RP5 family ran as normal with no chaotic peaks. Experiment undertaken by Sanger Centre, Cambridge.

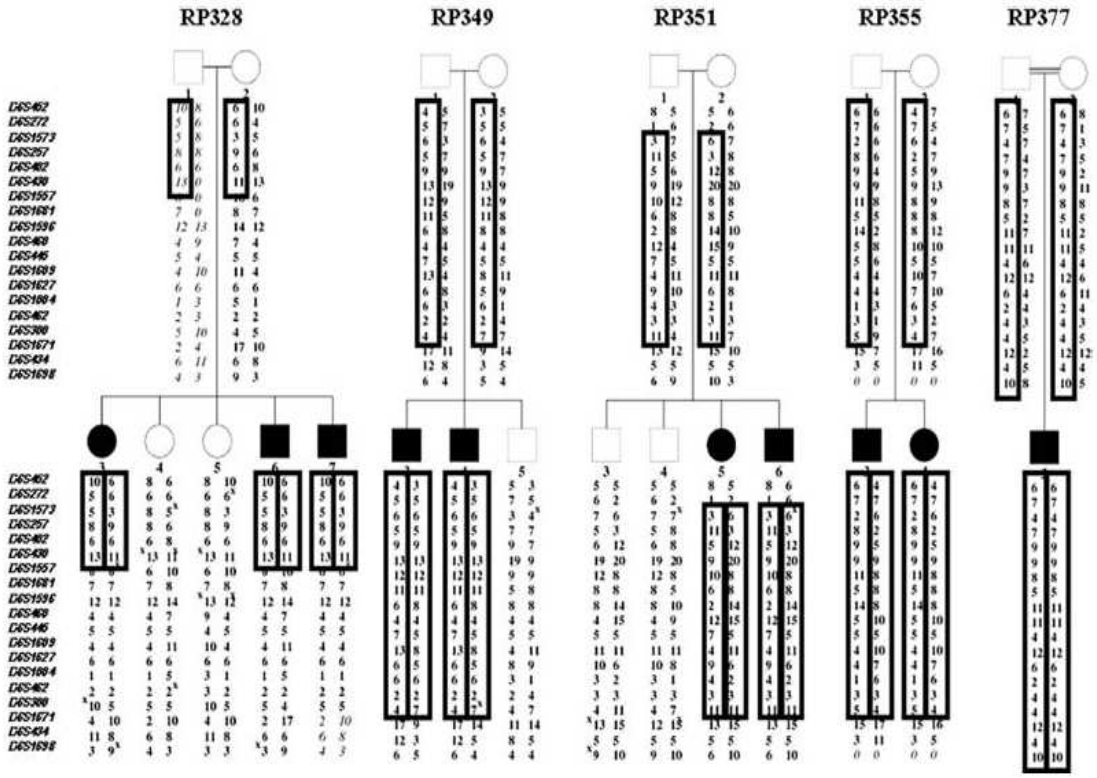


Figure 3.12 Haplotypes of the newly recruited Spanish *RP25* families, RP328, RP349, RP351, RP355 and RP377. Microsatellite markers on chromosome 6p12.1-q12 were used in this instance. Recombination events between markers D6S272 and D6S1557 occurred in families RP351 and RP328 respectively (Barragan *et al.,* 2008).

Table 3.3 Summary of the genes in interval B

Genes in red have previously been screened for mutations

Gene Annotation	Location (Mb)	Gene size/Exon number	Expression info	Published info on gene
*Q5T669*_human Fragment gene	65.38-65.39	8.87 Kb/2	None	Contains EGF-like domain.
Q5T1H1 Fragment gene	65.57-65.58	9.55Kb/3	Expression identified in tissues including eye, brain, lung and blood	Similar to Neurogenic locus Notch protein precursor.
Q9H557_human Fragment gene	65.65-65.67	26.05Kb/4	None	Contains EGF-like domain.
*Q5TEL3*_human Fragment gene	65.76-65.76	0.123Kb/1	None	Contains EGF-like domain.
*Q5TEL4*_human Fragment gene	65.77-65.77	0.075Kb/1	None	Contains EGF-like domain.
*Q5VVG4*_human Fragment gene	65.82-65.82	0.111Kb/1	None	Contains EGF-like domain.
Q5T3C8 Fragment gene	66.062-66.063	0.255Kb/1	None	Contains EGF-like domain.
EGFL11 EGF-like domain, multiple 11	66.10-66.26	161.09Kb/9	Expressed in several tissue lines, but mainly in eye and ovary.	Contains EGF-like domains
BAI3 Brain specific angiogenesis Inhibitor 3	69.40-70.15	750.22Kb/30	Ubiquitously expressed	BAI3 may participate in the early phases of ischemia-induced brain angiogenesis and in brain tumor progression. (Kee *et al.,* 2002).
LMPRD1/C6orf209 LMBR1 domain containing 1	70.44-70.56	21.03Kb/16	None	None
COL19A1 Collagen, type XIX, alpha 1	70.63-70.97	345.71Kb/51	Expressed in several tissue lines but predominantly blood.	Unlike all other tissues, expression of Col19a1 in the central nervous system gradually increases after birth. (Sumiyoshi *et al*., 2001).

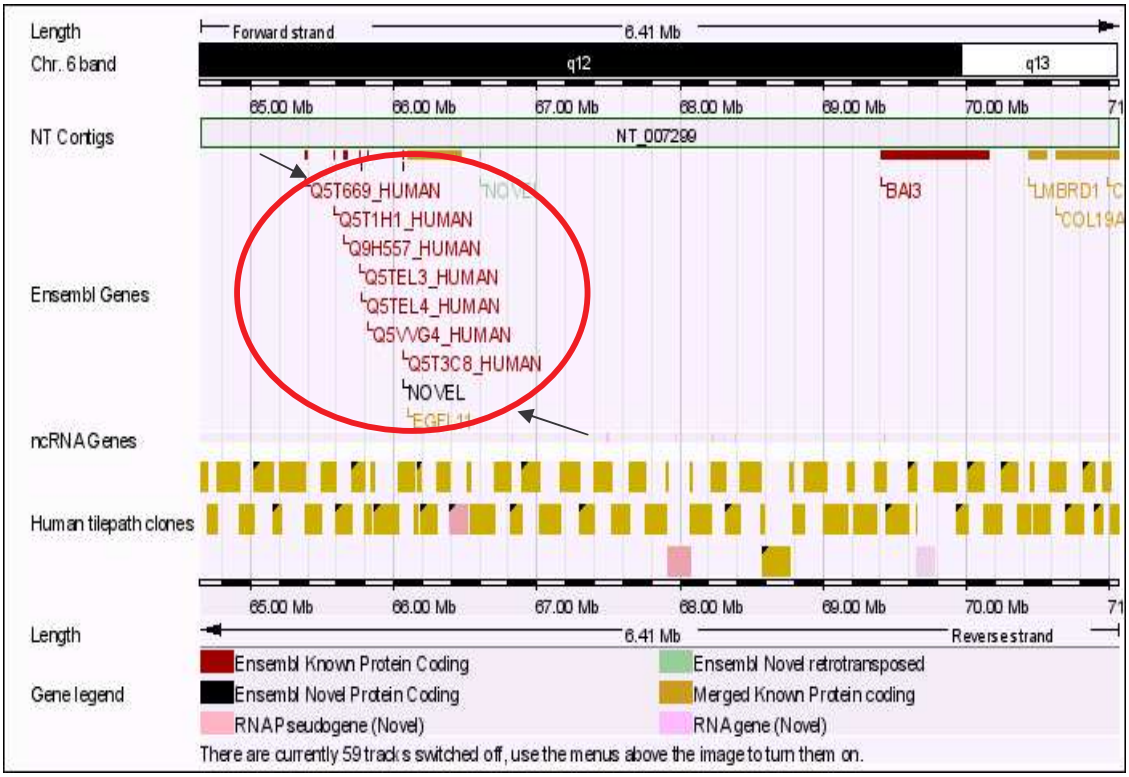
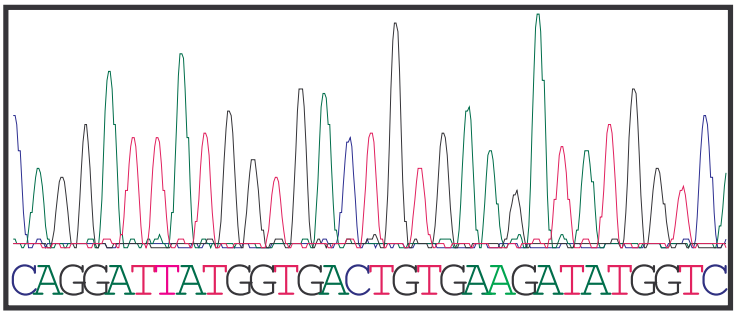


Figure 3.13 Ensembl view (version 43: April 2007) of the arrangement of the gene transcripts being sequenced. The common feature of all of the transcripts in the interval is an epidermal growth factor like domains.

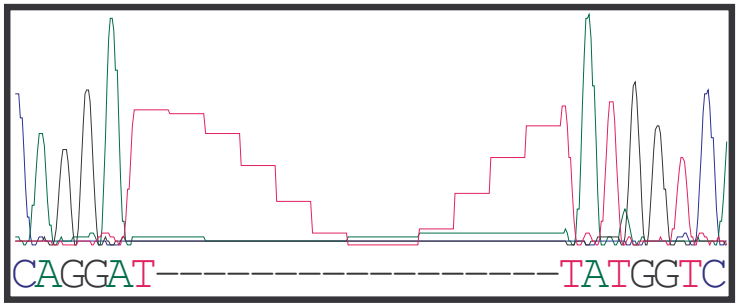
3.5.1 IDENTIFICATION OF A SECOND DELETION

While screening the four exons of *Q9H557_human*, a novel change was identified. This was a 17 base pair frameshift deletion in exon 2 and was detected in both affected members of family RP214. This deletion was not identified in the unaffected member of the family (Figure 3.14) and was also absent in any of the other Spanish RP families screened. This frameshift deletion causes a change in the reading frame of the amino acids and would ultimately lead to a truncated protein.

To further prove the phenotypic segregation of this 17 base pair deletion in family RP214, a restriction enzyme digest was performed on all members of the family. The digest was performed using the restriction enzyme *Tsp45I* (5'...GTSAC...3') (Promega, UK). Exon two of *Q9H557* was digested with 1 unit of *TSP451* and products were analysed on a 2% agarose gel. Segregation was revealed in this family as none of the affected members of RP214 demonstrated cleavage of the enzyme (Figure 3.15). The restriction digest was also performed on a control Spanish population of 96 members, to verify this is a novel, possible disease causing mutation (Figure 3.16). All members of the panel, 80 of which are demonstrated in figure 3.16 segregate normally when cleaved with restriction enzyme *Tsp45I*, demonstrating that this deletion is unique and is absent from the normal Spanish control population.



RP214 Parent



RP214 Affected

Figure 3.14 PCR amplification of exon 2 of *Q9H557*. **A** RP214 parent with normal coding transcript. **B** RP214 affected member with a 17 base pair deletion in exon 2 leading to a frameshift deletion that may lead to a truncated protein.

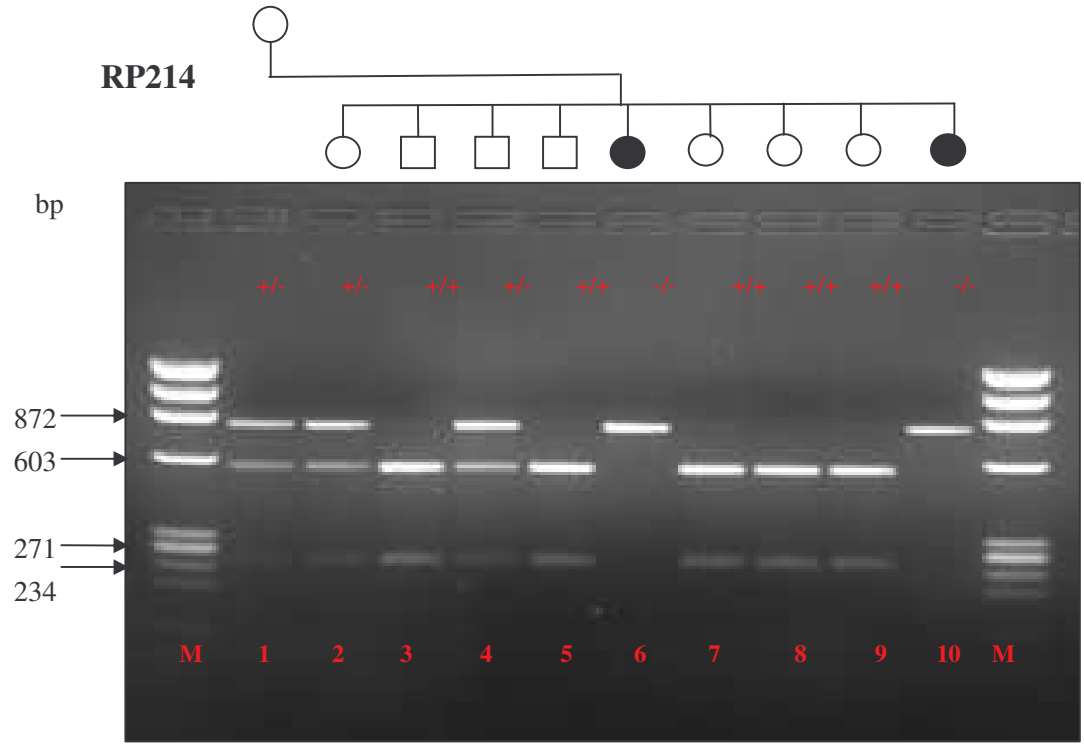


Figure 3.15 Restriction enzyme digest performed on all members of family RP214, using restriction enzyme *Tsp451* in the 17 base pair region of exon 2 of *Q9H557*. Affected members (6 and 10) have a deletion where the restriction enzyme cuts, with neither allele been cleaved by the enzyme (855bp). Carriers for the mutation (1, 2, and 4) have one allele cleaved with this enzyme, as they have one wildtype allele (600bp and 272bp) and one affected allele (855bp). Non-affected members (3, 5, 7, 8 and 9) have both alleles cleaved (600bp and 272bp). Segregation of this deletion is in accordance with the 17 bp deletion identified by sequencing.

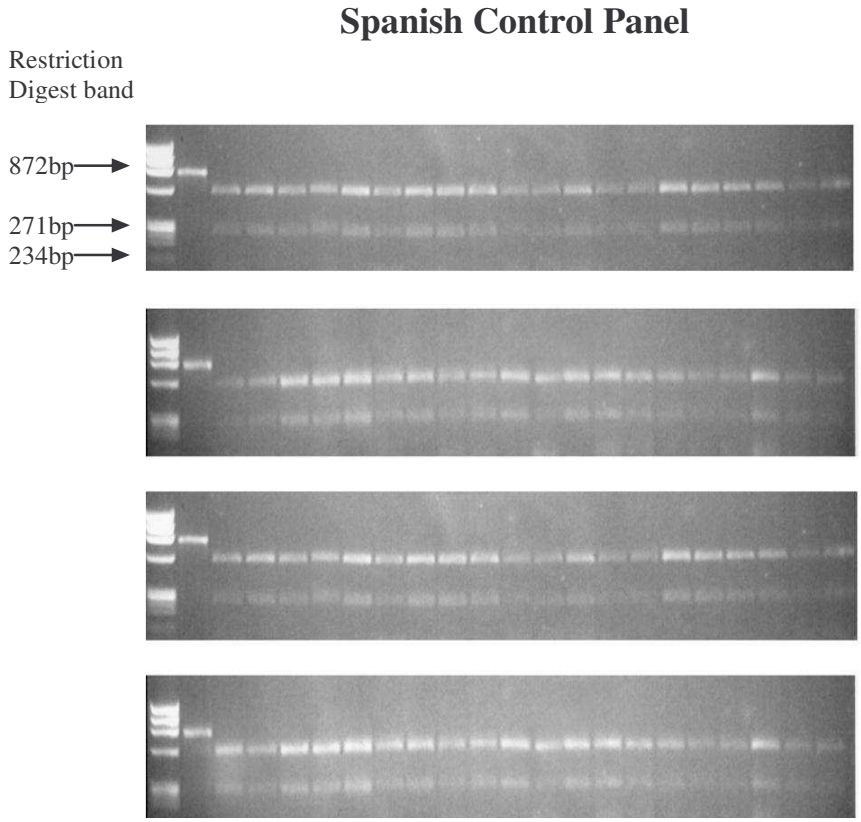


Figure 3.16 Restriction digest using restriction enzyme *Tsp451*, displaying 80 members of the 96 Spanish control population. Marker represents bands within a 1 Kb size range; lane 1 in each row is an affected member of RP5 family (872bp) where there is no digestion. All control members amplified normally and all were cleaved with the enzyme.

3.6 Extension of *RP25* transcript

To reiterate, two independent deletions in two unrelated RP affected families were identified in a region where a cluster of seven incomplete gene transcripts map to. As the 17bp deletion was identified in *Q9H557* and this is an incomplete transcript, a methodical bioinformatics investigation was performed to identify if further exons of this transcript existed. Using bioinformatics databases such as Ensembl (http://www.ensembl.org/index.html), the predicted size of this gene transcript increased from four to 14 exons. Eight of these newly predicted exons correlated to the exons of previously annotated transcripts which include, downstream of *Q9H557*:

Q5T3C8: one exon (66,062,478-66,062,732)

Q5VVG4: one exon (65,824,229-65,824,339)

Q5TEL4: one exon (65,771,073-65,771,147)

Upstream of *Q9H557*

Q5T1H1: three exons (65,579,997-65,589,548)

Q5T669: two exons (65,384,020-65,392,891) (Figure 3.12)

Two novel exons were also predicted as being part of this 14 exon novel transcript and the exons of *Q9H557* (1-4) were renumbered as exons 7-10. Using 5` and 3` RACE PCR, primers were designed in these exons to "walk" from exon to exon using human retinal cDNA (Clontech, France), a method also known as exon collection. Primers were designed flanking all 14 exons and PCR amplification was performed using

genomic DNA of all affected families and amplification of the full-length gene was possible. It is imperative to highlight that no translation start and stop codons were identified, and further RACE and bioinformatics predictions were required to fully characterise this gene.

3.6.1 Deletion in RP5 family

Once the extended transcript of *Q9H557* was identified, PCR amplification of all members of the affected families was performed. From this amplification it was possible to correlate with which exons the deletion in the affected members of family RP5 occurred in. Amplification of the affected members of RP5 failed for exons 6 through to 10 (previously exon 1 of *Q5TEL4* and exon 1-4 of *Q9H557*) of this newly predicted 14 exon transcript. Exon amplification was successful in the unaffected members of this family (Figure 3.17) supporting the phenotypic segregation of this mutation in the RP5 family. This result reaffirms the deletion identified through the CGH array on the RP5 family, however to reveal the breakpoints of this deletion further MLPA assay analysis would be necessary.

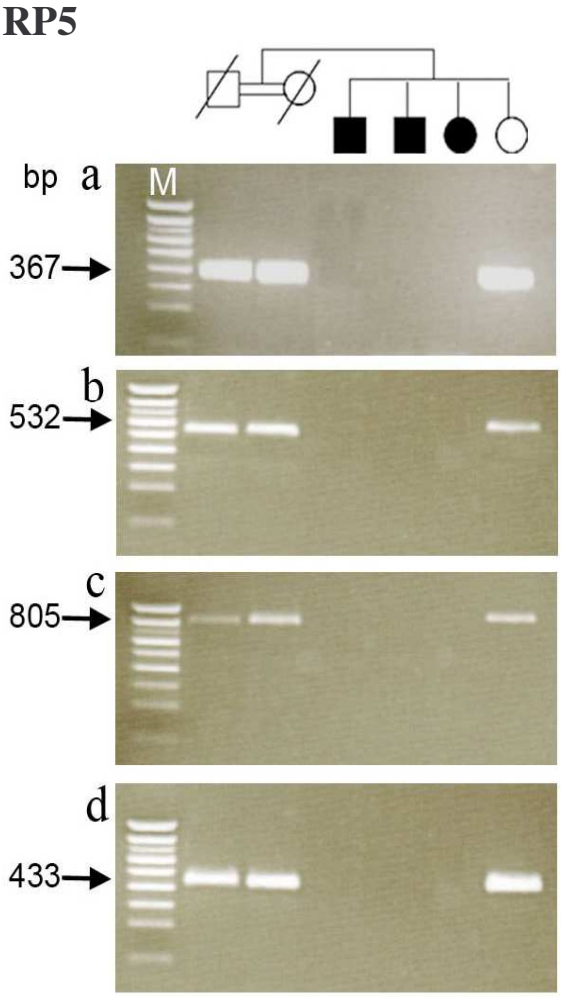


Figure 3.17 Amplification of exons 6-10 of RP5 members of the extended gene for RP25. a) Corresponds to exon 6 (previously exon 1 of *Q5TEL3*) which is 367 bp in size. b) Represents exon 7 (previously exon 1 of *Q9H557*) which is 532 bp in size. c) Corresponds to exon 8 and 9 together (previously exon 2 and 3 of *Q9H557*) (805bp), and d) corresponds to exon 10 (previously exon 4 of *Q9H557*) (433bp). Please note amplification failure of all 5 exons in each of the three affected family members.

3.7 Conclusion of Chapter 3

A functional candidate gene approach was undertaken to select genes for mutation screening in RP25 affected families. Highly efficient genetic databases were utilised to identify good candidate genes. 54 genes had been previously collectively screened and excluded as possible causative genes. Six genes were selected as possible candidate genes and were subsequently excluded as possible disease causing genes for RP25. Due to heterozygous SNPs being identified in a stretch of homozygosity in gene *PRIM2A* in affected members of two families (RP5 and RP214), a CNV study was performed. From the array analysis undertaken by collaborators in Sanger Centre, Cambridge, a deleted clone was identified in the RP25 interval in the RP5 family.

Logical and intensive mutation screening of genes in the RP25 interval in parallel with CGH array analysis on all members of an affected family revealed the causative *RP25* gene. Initial genome tilepath results revealed a deletion of a clone in the RP25 region of the affected members of the RP5 family. Based on the linkage of a further five Spanish families to the locus, the RP25 interval was narrowed down to 2.67 cM. Gene transcripts in region B of RP25 then became the focus of mutation screening as this is where the deleted clone mapped to. Seven incomplete transcripts (*Q5T669*, *Q5T1H1, Q9H557_human, Q5TEL3_human, Q5TEL4_human, Q5VVG4_human,* and *Q5T3C8*) were all screened for possible mutations.

One of these transcripts, *Q9H557* which has 4 exons, revealed a 17 base pair deletion in affected members of the RP214 family. A restriction enzyme digest concluded the segregation of this deletion in the family. A bioinformatics study determined that this gene *Q9H557*, along with the surrounding gene transcripts may all belong to one larger gene. 5´ and 3´ RACE and PCR amplification confirmed that these transcripts do belong to one larger gene; we postulated that this gene is 14 exons in size, identifying 2 novel exons confirmed through PCR amplification.

It is imperative to highlight that no translation initiation or termination site had been identified therefore suggesting this gene is larger in size. To verify the validity of the CGH clone deletion, a PCR analysis was performed on all members of the RP5 family using primers designed across the 14 exon transcript. The deletion corresponded to 5 exons of this 14 exon transcript and segregated in the RP5 family; however the breakpoints of this deletion had yet to be identified.

Chapter 4

Characterisation and Mutation Screening of *RP25* gene: *EYS*

4.1 Identification of full length *RP25* gene

Following the preliminary identification of the extended *RP25* transcript, subsequent experiments were performed to identify the full length *RP25* gene.

4.1.1 Identification of new exons of *RP25* transcript

In Chapter 3 it was established that the *RP25* transcript spans 14 exons, however the transcription start site and translation initiation and stop codons remained to be determined. Through further RACE experiments and intensive bioinformatics analysis (with the assistance of Professor Chris Ponting and Dr. Leo Goodstadt, University of Oxford), it was possible to generate primers to sequence the coding region bi-directionally of the 14 exon transcript.

Initially, first-strand retinal cDNA synthesis was performed, prior to 5´ and 3´-RACE PCR analysis. Gene specific primers (GSPs) were primarily designed in exons 1 and 4 of *Q9H557* (exon 7-10 of the 14 exon transcript). Subsequently, several GSPs were designed upstream and downstream of the first primer pair, once a novel exon was

identified, new primers were designed for the adjacent exon and a chromosome (or transcript) walking technique was employed using human retinal cDNA. Sequence analysis of the various PCR products was then performed bi-directionally, followed by alignment of the multiple contigs to identify the transcript structure of this novel gene (Figure 4.1). Interestingly, it was determined that *EGFL11*, a previously characterised and annotated gene formed part of this novel transcript.

The *EGFL11* segment of the *RP25* gene formed the 5´ end; with the translation start site positioned in exon 4, as exons 1-3 were identified as non-coding. Subsequently, incorporating the bioinformatics predictions of exons, further internal exons were identified including the previously annotated *Q5T1H1*. Consequently, the full length gene increased in size from the initial 4 exons of *Q9H557*, to 26 exons, which incorporated the annotated transcripts, *Q5VVG4, Q5TEL3* and *EGFL11* (Figure 4.2). Furthermore, analysis of gene product predicted that several EGF-like repeat domains exist in the corresponding protein (Figure 4.2). However, it is imperative to highlight that though the translation start site had been identified, the stop codon was yet to be established, with further RACE and bioinformatics analysis to be performed.

Figure 4.1 Overview of RACE PCR reactions depicted on a 1% agarose gel. M represents the 10 kb marker used. 1-5 depict amplicons of varying size. The stepwise size of each band illustrates the increased size of each amplified fragment.
1-ex2R2 of 14 exon *RP25* transcript and *EGFL11* ex8F (expected amplicon size of 600bp)
2-ex2R1 of 14 exon *RP25* transcript and *EGFL11* ex6F (900bp)
3- ex2R2 of 14 exon *RP25* transcript and *EGFL11* ex4aF (1,170bp)
4- ex5R of 14 exon *RP25* transcript and *EGFL11* ex2aF (1,944bp)
5-Control-ex7F1 and ex10R of *RP25* transcript (600bp).

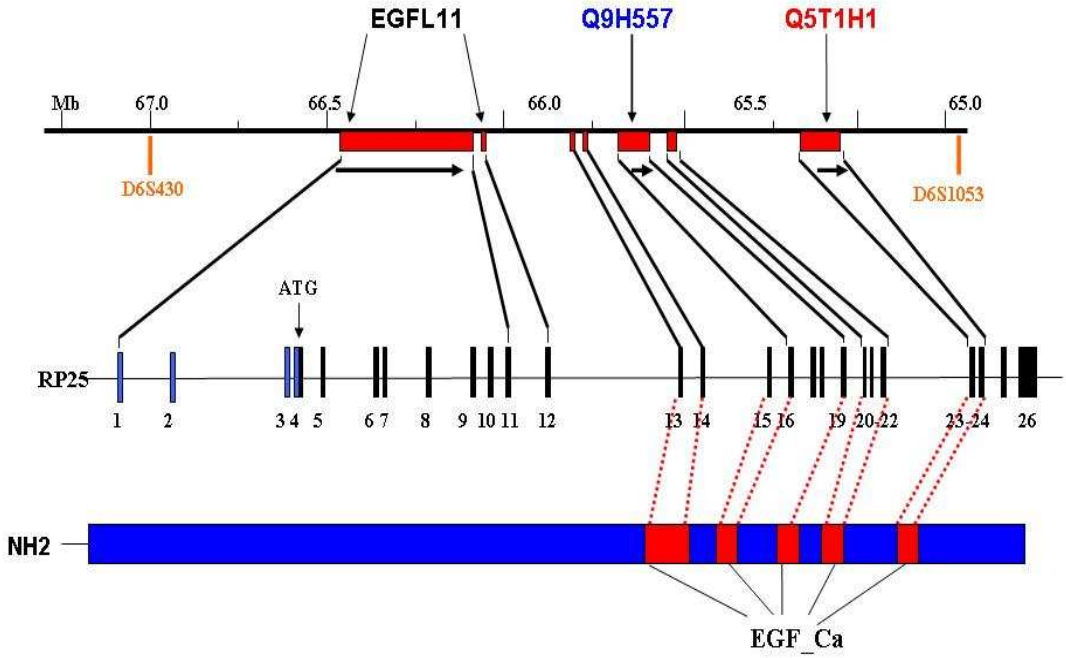


Figure 4.2 Initial prediction of the *RP25* gene, the transcript spans 26 exons, including 9 exons from the previously characterised gene *EGFL11*. Other exons correspond to previously predicted transcripts of *Q9H557* and *Q5T1H1*. Also depicted is the predicted EGF like domain structure of the gene.

4.1.2 Full length structure and size of gene

Once the final rounds of 3´ RACE were complete, determining the full length gene sequence was possible (Table 4.1 intron-exon boundaries). The gene encompasses 30 exons from 9 previously annotated genes including; *EGFL11, Q5T3C8, Q5VVG4, Q5TEL3, Q9H557, Q5T1H1, Q5T669, ENSG00000214559,* and *Q9NQ15* (Figure 4.3). The overall coding region of *RP25* is approximately 9 Kb in size and contains 3,145 amino acids (see supplementary data). The gene consists of 40 coding exons, and three non-coding exons. The translation start codon is in exon 4 and translation termination codon is at the end of exon 43. *RP25* maps on chromosome 6q, spanning 66,473,990-64,487,835 Mb. The size of the gene spans just over 2 Mb, making it one of the largest genes in the human genome, the largest gene being dystrophin which is 2.5Mb in size (Roberts *et al.,* 1993).

One significant outcome of these RACE experiments was the identification of the absence of a 1,238 bp segment spanning the 3´ end of exon 29 from the human online genomic databases. This reaffirms that human genome databases, although highly accurate and reliable, still require traditional molecular techniques to identify novel sequences. This 1,238-bp fragment did not alter the reading frame of the previously identified coding region.

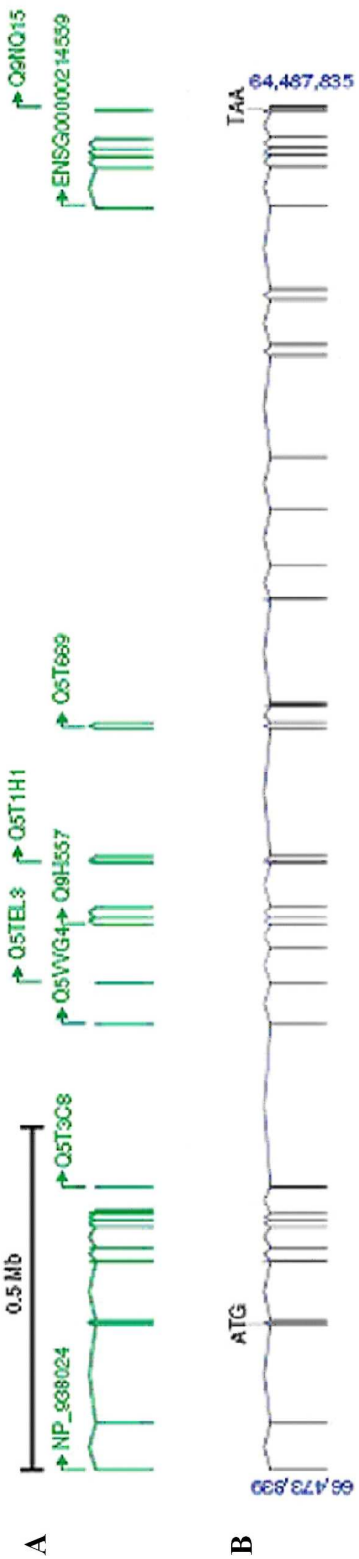


Figure 4.3: A- Corresponding gene transcripts of *RP25* gene. **B-** An overview of all 43 exons in the gene, from the translation start site which is in exon 4 to the stop codon which lies in exon 43. The gene spans from 66,473,990-64,487,835, spanning 2 kb in size, making it one of the largest genes yet identified in the human genome.

Table 4.1 Exon and Intron Sizes of Human *EYS*
(See also supplementary data)

Exon/Intron	Exon size (nt)	Intron Size
1	43	67,243
2	115	143,785
3	135	252
4	945	3,956
5	114	85,227
6	194	2,569
7	128	17,978
8	115	30,769
9	160	9,281
10	140	8,892
11	167	38,861
12	257	238,136
13	114	59,911
14	122	51,668
15	122	33,050
16	260	9,984
17	97	184
18	108	15,271
19	146	63,875
20	172	934
21	79	8,068
22	200	187,133
23	125	8,588
24	116	24,109
25	193	1,128
26	1,767	150,871
27	191	2,897
28	92	47,334
29	151	80,403
30	113	76,146
31	233	148,590
32	147	15,365
33	154	67,155
34	109	14,472
35	221	120,025
36	173	57,814
37	183	16,966
38	167	809
39	145	9,925
40	175	15,373
41	173	35,719
42	162	4,719
43	*1,689*	
***EYS* total size**		**1,985,847**

4.2 cDNA Expression and RT-PCR

4.2.1 cDNA expression profile

Tissue expression analysis is a fundamental experiment and was performed to identify in which tissue types the *RP25* gene is expressed. cDNA samples from human retina, brain, kidney, liver, heart, skeletal muscle, pancreas, lung and placenta (Quick-Clone; Clontech) along with lymphoblast, Y79, ARPE19 and HELA cell lines were all incorporated into the expression profile experiment. Y79 cells are a retinoblastoma cell line, representative of photoreceptor cells, retinal pigment epithelium (RPE) cells are represented by the ARPE19 cell line, while HELA cells are an immortal cell line. cDNA specific primers were designed in exon 4 (forward) and exon 11 (reverse) (see appendix). The amplicon of this fragment was 1.8 Kb in size.

Exons of the ubiquitously expressed gene, *PGM1*, (phosphoglucomutase 1) (Whitehouse *et al*., 1992) were amplified in parallel as a positive control (Figure 4.4). PCR products were electrophoresed on a 2% agarose gel and visualised through ethidium bromide staining. The RT-PCR results for the retina tissue (Figure 4.4) demonstrates a distinct amplicon that is intensely stained suggesting that the *RP25* transcript is retina specific. Expression of the fragment was also detected in the photoreceptor-like Y79 cells. Amplification in all other tissues was unachievable. The *PGM1* control amplification verified the high quality of all tissue and cell cDNAs

used. A further RT-PCR was performed using primers designed in exon 26(F) and exon 40 (R), with the same result being produced. It is possible that smaller transcripts, upstream or downstream of exon 4-40 may exist in other tissues; however it can be deduced that the full length transcript of *RP25* is retinal specific. This is a very significant result, as it can be concluded that the *RP25* gene is the largest eye specific gene identified in the human genome to date.

4.2.2 Full length RT-PCR

To verify the predicted size of the coding transcript of *RP25* as being accurate, a reverse transcriptase PCR (RT-PCR) experiment was performed. By reverse transcribing the RNA strand into its DNA complement using reverse transcriptase, the resulting cDNA is amplified by PCR. Two cDNA specific primer pairs were designed to PCR amplify the coding transcript in two overlapping fragments. The first primer pair was designed from exon 4 to exon 26 (4.8 Kb), the second was designed to amplify a cDNA fragment from exon 26 to exon 43 (4.1 Kb) (see appendix) with the overlap in exon 26.The successful amplification of these two fragments confirms that the retinal transcript of the *RP25* gene is the predicted size of 9.1 Kb and 3,145 amino acids.

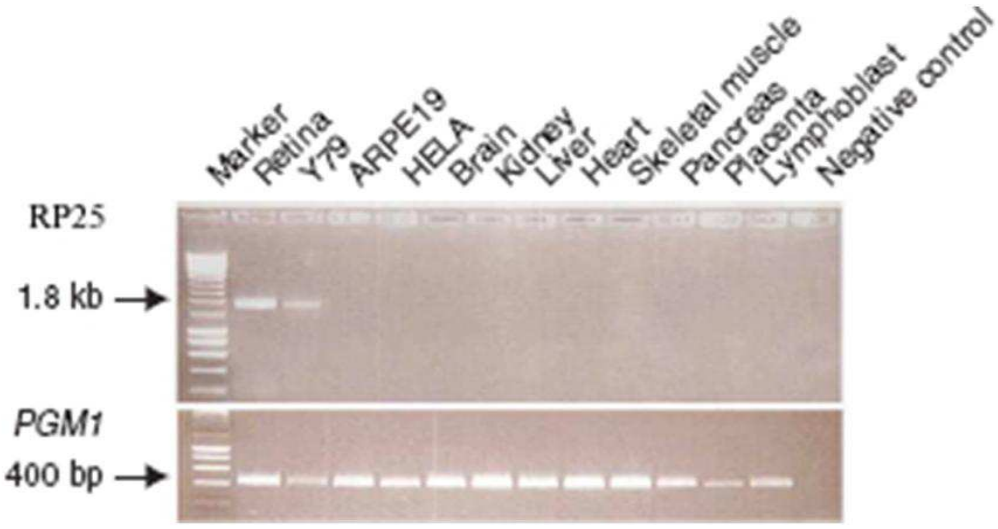


Figure 4.4 A cDNA expression profile for the full length gene of RP25. Primers spanning 8 exons (exon 4-11) were used to amplify a large coding region. The cDNA amplicon was 1.8 kb in size. A retinal specific cell type (Y79) was incorporated in this expression experiment. A band specific for 1.8 kb was amplified in retinal cDNA and in the Y79 cell line. The control experiment was performed using *PGM1* primers, which is a ubiquitously expressed gene.

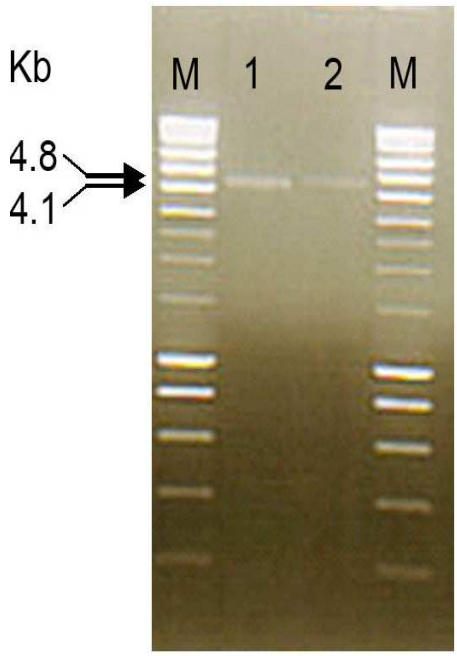


Figure 4.5 A RT-PCR experiment performed on two overlapping fragments of the coding transcript of the *RP25* gene. The first fragment ran from exon 4 to exon 27 this was 4.1kb in size. The second fragment incorporated exon 27 to exon 43. The marker used in this experiment was a 10 kb ladder. The full length of the transcript stands at 9.1kb. Thus the two overlapping fragments correspond to the full length size of the cDNA transcript.

4.3 Protein structure and domains

Protein structure prediction was carried out with the expertise of Professor Chris Ponting and Dr Leo Goodstadt of Oxford University. The domain architecture of *RP25* was predicted using the SMART (Simple Modular Architecture Research Tool) protein domain database server. It allows the identification and annotation of genetically mobile domains and the analysis of domain architectures User interfaces to this database allow searches for proteins containing specific combinations of domains in defined taxa. EGF, Ca^{2+}-binding EGF and Laminin-G domains were identified as the combination of domains in the RP25 protein. EGF-like and Laminin G domains are commonly found in extracellular proteins. Many proteins containing Laminin G domains are involved in a variety of roles, including signaling, migration and assembly. Laminin-1 appears to be a key molecule in early embryonic basement membrane assembly (Sasaki *et al.,* 2004). EGF-like domains are known to present in a more or less conserved form. Proteins with EGF-like domains are often quite large (> 1000 amino acids) and have multiple copies of this domain in addition to other domains known to be involved in specific protein-protein interactions (Nakayama *et al.,* 1998). The EGF, Ca^{2+}-binding EGF and Laminin-G domain predictions were further refined manually using seed alignments from SMART and hmmsearch from the HMMER suite of programs. Functional EGF domains were found by conservation of the consensus cysteines required for disulphide bond formation. EGF- Ca^{2+} domains were identified if they retain consensus negatively-charged residues.

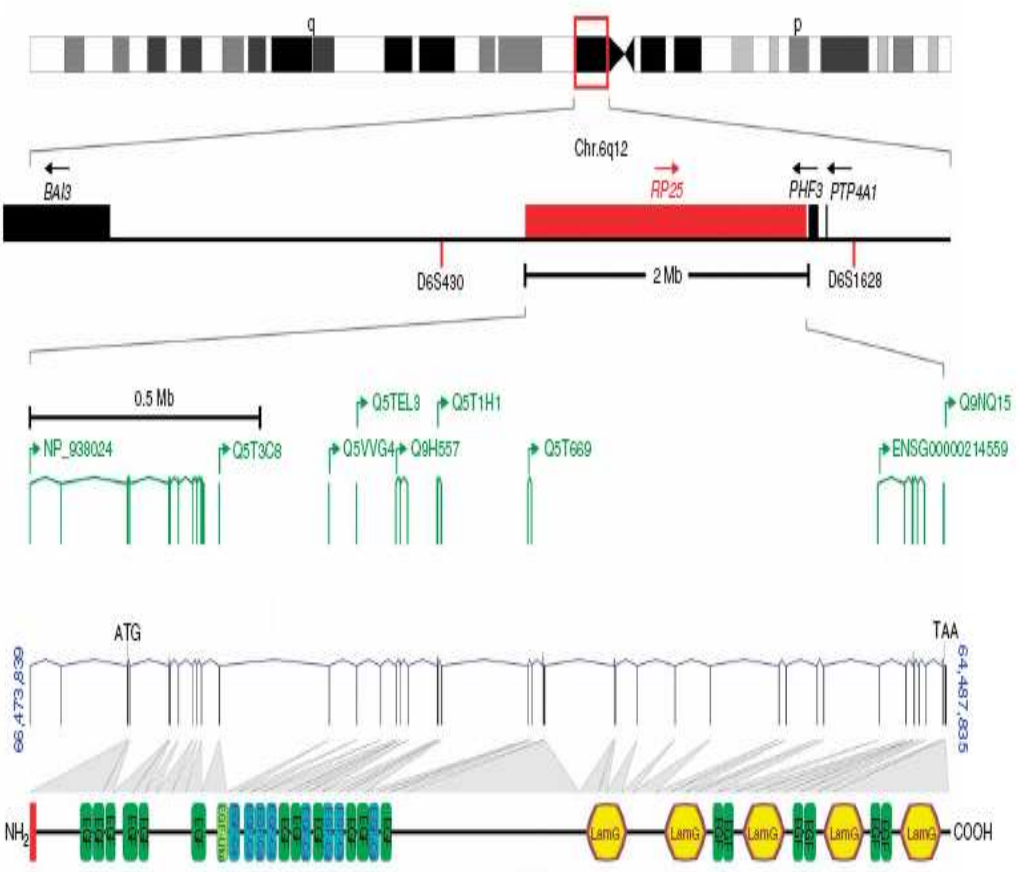


Figure 4.6 Overview of the size and structure of the *RP25* gene. *RP25* is mapped to chromosome 6q12. The genomic size of the gene is over 2 Kb in length. The corresponding previously identified transcripts within the *RP25* gene with the localisation of exons spanning the *RP25* gene. The corresponding domains and peptide signal (red) in the *RP25* protein, including EGFL, LaminiG and Ca^{2+} domains is demonstrated at the bottom of the figure .

4.4 Homolog of human *RP25* in *Drosophila melanogaster*

Once the domain structure of the RP25 protein was ascertained identity alignments to other characterised proteins was performed. As previously depicted, the featured regions of the RP25 protein include EGF like, Ca^{2+}- binding EGF and Laminin (LamG) domains. Based on proteins previously identified with EGF-like domain repeats, the repeat pattern of cysteines in the EGF-like domains of the RP25 protein may be an essential factor in its overall tertiary structure. The trademark of extracellular proteins as Notch expressed in *Drosophila melanogaster* (*D. melanogaster*) is a run of such EGF-like domain repeats (Weinmaster *et al.,* 1991). LamG is thought to have a diverse role in cell adhesion, signalling, migration, assembly and differentiation (Beckmann *et al.,* 1998). BLASTP was applied to compare the RP25 protein against all other available sequenced species. Each species was aligned against the human protein, through which an interesting alignment to the *D. melanogaster* (fruit fly) genome was determined. All *D. melanogaster* and human orthologs whose sequences contained predicted LamG domains separated by one or two EGF domains were selected. The LamG domain sequences were manually aligned and the phylogenetic distances between them calculated using protdist from PHYLIP. An ortholog of the human *RP25* gene was subsequently discovered in *D. melanogaster*.

4.5 *RP25* Gene annotation: *EYS*/(SPAM/SPACEMAKER)

The N terminus fragment of the *RP25* gene (*EGFL11*_human) was not identified as having a homolog in *D. melanogaster*; however the rest of the gene was established to be highly homologous to a previously predicted gene. The ortholog of interest is called *eys* (*eys* shut) with the protein name of spacemaker/spam. The domain organisation of *eys* is comparable to that of the human homolog protein, with 14 EGF-like and four laminin G-like domains positioned in a similar order (Figure 4.7). Based on this breakthrough, the *RP25* gene was given the name *EYS* and the protein name of SPAM.

Eys has previously been identified as a gene that is essential for the formation of matrix-filled inter-rhabdomeral space in the D. melanogaster eye (Husain et al., 2006). This inter-rhabdomeral space is an important aspect of the structure of the Drosophila eye, as it separates each rhabdomere (photoreceptor) from each other. The adult eye consists of approximately 800 hexagonal ommatidia or facets, each ommatidium is a structure which contains eight photoreceptor cells, and each associated with a photosensitive membrane called a rhabdomere. Photoreceptor cells one through six of each ommatidium is placed radically around cells seven and eight. Rhabdomeres of cells seven and eight occupy a central position, cell seven above cell eight (Figure 4.8). The visual system of the D. melanogaster is built following the

principle of neural superposition that is believed to have evolved from an apposition compound eye type which is the visual system of most insects (Husain et al., 2006).

4.6 SPAM orthologs in the eyes of other insects

Identification of an ortholog has been determined in *D. melanogaster*, but what about in other insect species? Orthologs of spam have been identified in the eye of other insect species with a similar rhabdom system as that in the *D. meanogaster* such as the housefly (*Musca domestica Linnaeus*). This system is called an "open rhabdom system" as each rhabdom is separated by an interrhabdomeral space; it is within this space that spam is expressed. Other species such as the honeybee (*Apis mellifera*), rust red flour beetle (*Tribolium castaneum*) and the mosquito (*Anopheles gambiae*) have an apposition visual system which is a subclass of the compound eye. In these, each ommatidium detects a different area in the visual field. All photoreceptor cells (PRCs) within one ommatidium collect light from the same area and their rhabdomeres are not required to be optically individual and instead tightly adhere to each other, which is also known as a closed or "fused rhabdom system". The *D. melanogaster* eye has evolved from an apposition eye in this "open rhabdom" system (Figure 4.9). Interestingly, orthologs of *D. melanogaster* spam have not been identified in the genomes of such insects as the honeybee, rust red flour beetle and the mosquito, in other terms, species with compound apposition eyes (closed rhabdom system). From this it can be concluded that spam is probably restricted to insects that acquire an open rhabdomeric visual system.

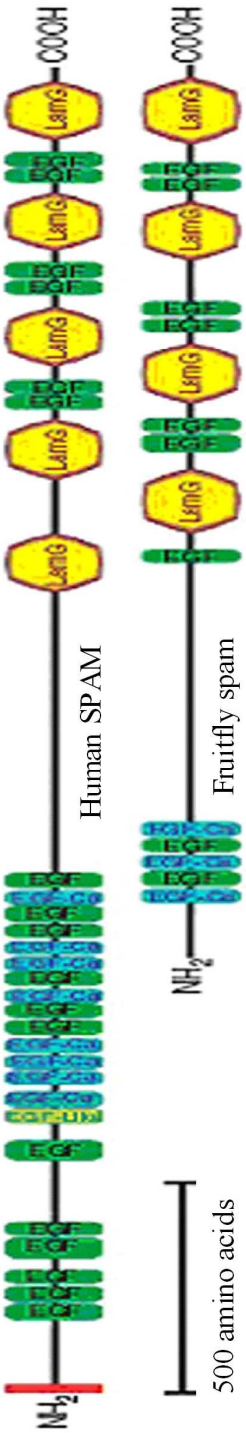


Figure 4.7: Alignment of the protein structure of human SPAM (*RP25*) and fruit fly spam. The first fragment of SPAM is not highly homologous in Drosophila, however the second fragment has high similarity between both species. The Laminin G and EGFL domains are highly similar in both protein structures.

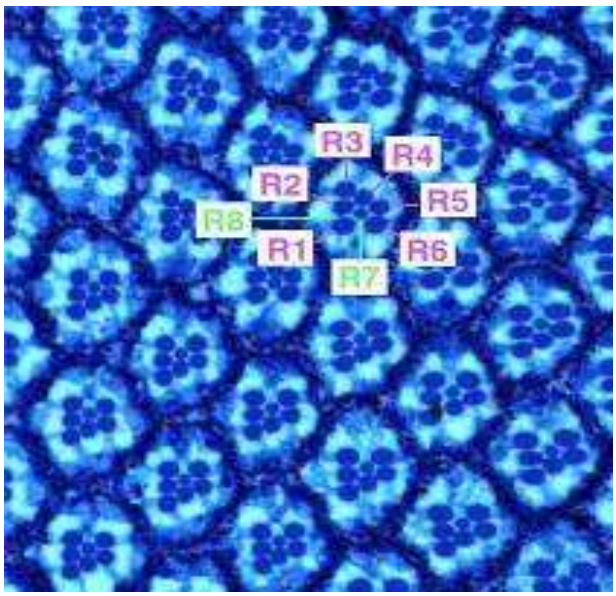


Figure 4.8 A depiction of a Drosophila compound eye. In each eye, there are approximately 800 ommatidia; in each ommatidium there are 8 rhabdomeres. Each rod shaped rhabdomere is representative of a photoreceptor cell.
(Taken from http://www.nimr.mrc.ac.uk/molneurobiol/salecker/eyesection/)

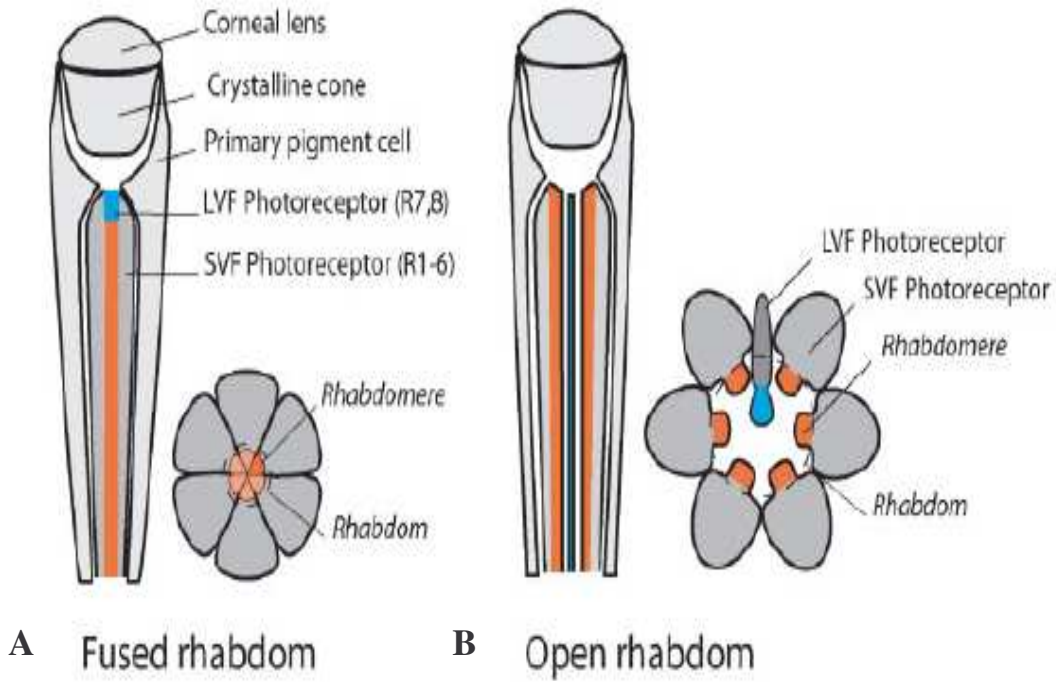


Figure 4.9 Two types of visual systems in insects. **A**- "Closed System"- The visual system of honey bees, in this system all rhabdomes are fused together. **B**- "Open System"- Evolved from the closed system, in this system all rhabdoms are separated from each other, this is the visual system of the fruit fly. LVF-Long visual fibre photoreceptor. SVF-Short visual fibre phororeceptor. R-Rhabdomere. Adapted from Osorio, 2007.

4.7 *EYS* and orthologs across species

In order to determine orthologs of *EYS* in other species, a BLAST search against all available sequenced genomes was performed. A phylogenetic tree was constructed by Dr. Leo Goodstadt, consisting of a selection of genomes across several phyla's. The phylogenetic tree (figure 4.10) was constructed through BLAST results of exon 43 of EYS. The mutual rate for each species used was then determined. The mutual rate is based on the redundant genetic code (d_n/d_s- number of non synonomous sites over the number of synonomous sites in the selected sequence). The mutual rate for conserved genes is usually approximately .1%. If this genetic code moves towards 1, the gene is becoming redundant in the species (Choi *et al.,* 2005). Resulting from the phylogentic tree analysis it was concluded that the *EYS* gene is found across many mammalian species including monotremes (platypus), marsupials (opossum) and chordata including the rabbit (*Oryctolagus_cuniculus*) and the pig (*Sus domestica*). It is also found with high identity in non-mammals such as the chicken (*Gallus gallus*) and in the zebrafish (*Danio rario*). However, interestingly this gene has somehow been disposed of with, in several species genomes (Figure 4.10) such species include the armadillo (*Dasypus novemcinctus*), little brown bat (*Myotis lucifugus*) and ruminants (cattle and sheep).

EYS also appears to be severely disrupted in rodents and has gone through more than 3 reading frame shift disruptions in three rodent species (mouse, rat and guinea pig). These frameshift disruptions lead to the pseudogenisation of *EYS* in these genomes. To confirm these predictions, BLAST was performed against these genomes using

EYS. Results of this experiment reinforced the phylogenetic tree data, that *EYS* has been disrupted in mice and rats. A short sequence run from *EYS* is depicted in figure 4.11. To further confirm the accuracy of these predictions, PCR amplification of exons belonging to *EYS* using mouse and rat retina cDNA. Mouse retina cDNA was prepared from RNA collected from mice bred in house. Rat RNA was purchased (Clontech, France) and cDNA was prepared in house. Amplification of a *GAPDH* exon (see appendix for primers) was incorporated into the experiment as a control and cDNA specific primers were designed across several exons of *EYS*, in exons 14, 26 and 43. Failure of amplification of these three amplicons spanning across *EYS* confirms that this gene structure is absent or has become a pseudogene, or is possibly assembled differently in mouse and rat genomes (Figure 4.12).

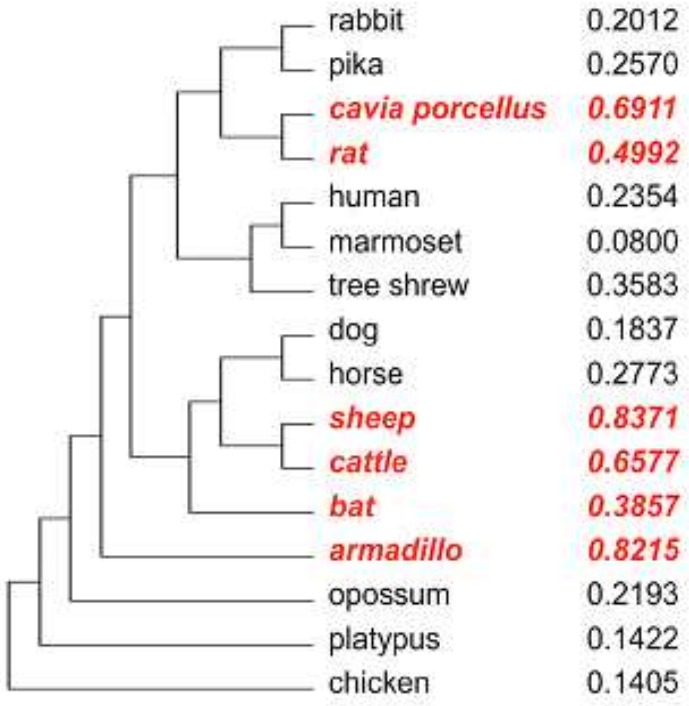


Figure 4.10 Phylogenetic tree depicting the independent disruption of *eys* (spam) in each of the guinea pig, rat, sheep, cattle, bat and armadillo lineages (in red). There are no traces of sequence correspondence to multiple regions of mouse genomic sequences in this region in the assembled genome which explains the absence of *eys* (spam) in the mouse. The final exon (43) of *EYS* was used to determine the mutual rate of eys across each species (numbers next to each species). This tree was constructed by Dr. Leo Goodstadt.

Species	Chromosome & nucleotide number position	Sequence
Homo_sapien	**6**:66416664 **6**:66416547	CATAGGGCCATGAAATGTCAATTGCAAATGAAATTGTATGATTTT-TCAGTTT-GTTTAGAAAAGATTTTGGCCATAAATCTAGAGATACTCATACTAGTTTTGTTG GAGGTGATGTTTA
Canis_familiaris	**12**:31905674 **12**:31905556	CATAGTGCTACAAAATGTCAATTATAAAATAAATCATATAATGTTCACACTTA CTTTTACAAAAGATGTTAATCATAAATCTGGAAATACTTAACCATGG-TTTTTG GAAGTGCCTTTTA
Equus_caballus	**20**:58311975 **20**:58311858	CATAGTGCAGTGAAATGTCAATTGTGAAAGAAACTGTATAATTGTCACACTTT-GTTTACAAAAGATGTTAACCATTAGTCTGGAAATACTTGAACCCAT–TTTTTG GAAGGGCCTTTTA
Mus_musculus	_____	…… ……
Rattus_norvegicus	_____	…… ……

Figure 4.11 Blast alignment of a short fragment of *EYS* against several species. This alignment reveals an obvious absence of *EYS* in *Mus musculus* (mouse) and *Rattus norvegicus* (rat). There is a high similarity of this particular sequence in *Canis familaris* (dog) and in *Equus cabullus* (domestic horse).

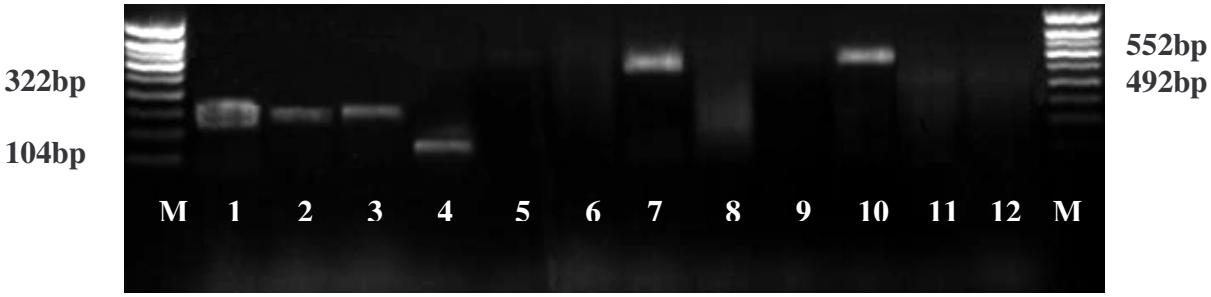


Figure 4.12 Amplification of *EYS* exons using human, mouse and rat retinal cDNA.**M-**1 Kb Smart fragment ladder**. 1**- Human *GAPDH* (exon 4-322 bp). **2**- Amplification of *GAPDH* exon 4 using mouse retinal cDNA. **3**- Amplification of *GAPDH* exon 4 using rat retinal cDNA. Lane **4** represents human cDNA amplification of exon 14 (104bp) of *EYS*. Lanes **5** and **6** demonstrate no amplification of exon 14 in mouse and rat retinal cDNA. Lane **7** is human exon specific amplification of exon 26 (492 bp), there is no amplification of mouse and rat retinal cDNA for this exon (**8** and **9**). The final three lanes (**10**, **11** and **12**) represent attempted amplification of exon 43a (552bp), only human cDNA (**10**) amplified successfully.

4.8 Mutation Screening of *EYS*

To summarise, two independent deletions have been identified in two unrelated Spanish RP families (RP5 and RP214). The two changes identified at this stage were a deletion of 100 Kb (which spans 5 exons) in family RP5, though the breakpoints of this deletion hadn't been defined and a 17 base pair deletion which was identified in the RP214 family.

Through exonic mutation sequencing, our colleagues in Spain identified a heterozygous null mutation in the RP73 family, this change was identified in exon 12 which leads to a stop codon (Figure 4.13). As RP25 is a recessive disorder, the requirement is for both alleles to have a mutation in order for a disease phenotype to prevail. Since this mutation was identified as a heterozygote, a further mutation is necessary to be identified for this change to be considered as disease causing. Exonic sequencing of the full length of *EYS* was performed, however no second mutation was identified, therefore a CGH array analysis was necessary for this family in order to identify if there is a second mutation, this possibly being a large deletion.

My colleagues identified two further unique mutations in two unrelated Spanish RP families. In family RP328, a homozygous point deletion of one base pair was identified in exon 28 leading to a truncated protein (Figure 4.14). A null mutation causing an amino acid change of tryptophan to a stop codon was identified in exon 41 of affected members of RP349 (Figure 4.15). In total, five unique mutations in five unrelated

Spanish RP families had been identified, however significantly a second change was necessary for the change in family RP73 to be considered as disease causing. All changes identified in the affected Spanish families are summarised in Table 4.2.

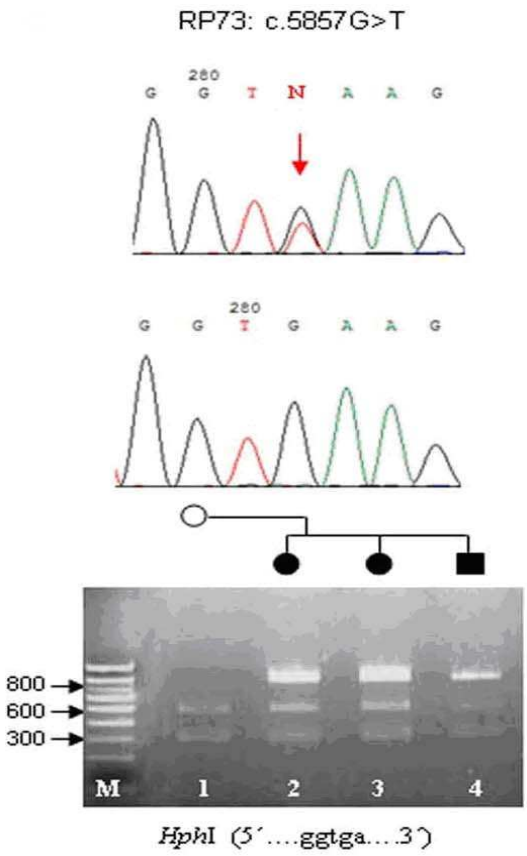


Figure 4.13 Chromatogram view of the heterozygous mutation c.5857 G>T in family RP73 and electrophoresis depiction of the change in the affected members of the family. A restriction digest using *Hph1* was performed to test the segregation of the change in this family. This experiment was performed by collaborators in Seville, Spain.

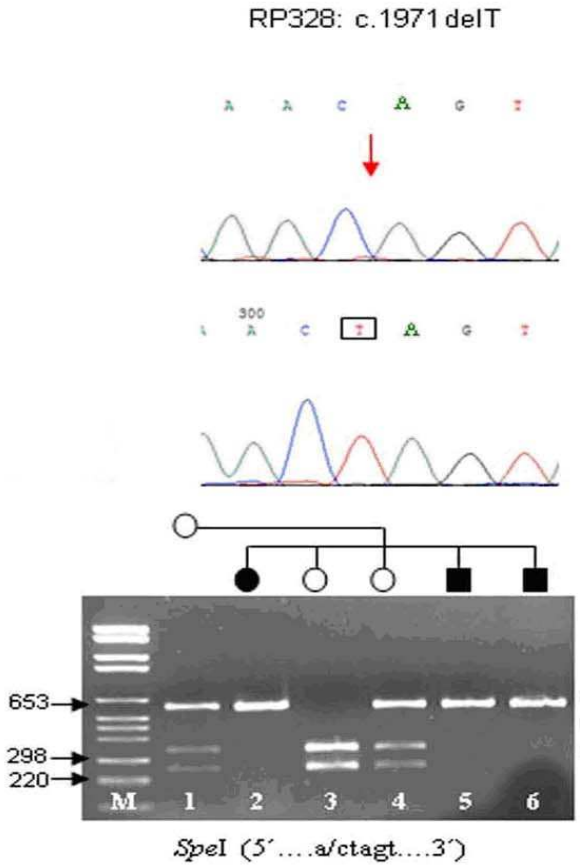


Figure 4.14 A chromatogram and electrophoresis view of the single bp deletion, c. 1971delT in family RP328. A restriction digest (*Spe*I) was performed on the family to demonstrate the segregation of the deletion in all members of the affected family. This experiment was performed by collaborators in Seville, Spain.

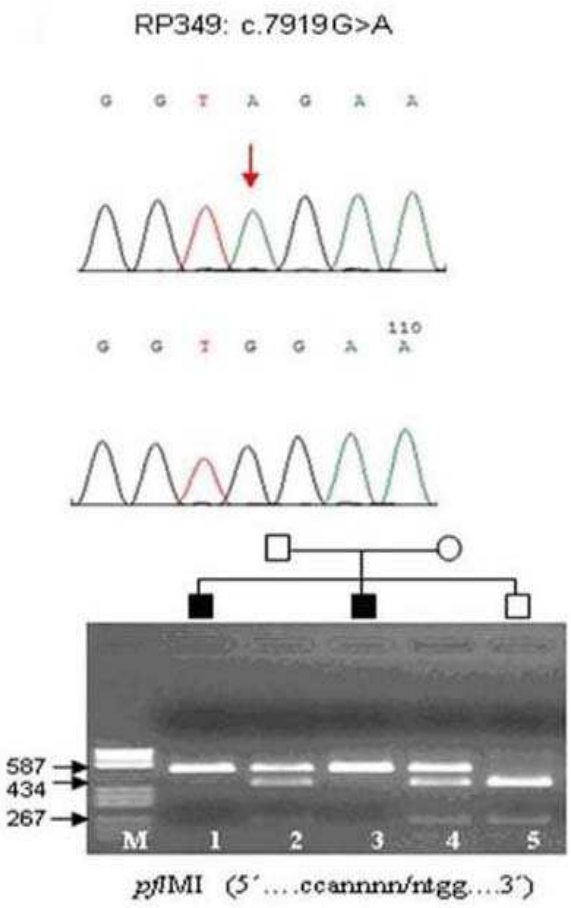


Figure 4.15 A chromatogram and electrophoresis view of the homozygous base pair change c.7919G>A in the RP349 family. A restriction digest using enzyme *pfl*MI was performed on all members of the affected family to demonstrate the segregation of the mutation. This experiment was performed by collaborators in Seville, Spain.

Table 4.2 Summary of Mutations identified in *EYS* in arRP Spanish families. (Table from Abd-El Aziz *et al.*, 2008)

Family ID	Exon	Nucleotide position	Protein alteration	Type of mutation
RP214	17	2710_2726del17	D904QfsX17	Homozygous
RP5	15–19	2260-51191_2992+45990	S754AfsX6	Homozygous
RP328	12	1971delT	T657TfsX5	Homozygous
RP73	12	[1767-24596_2023+238135]	R589RfsX5	Heterozygous
		+		
	28	[5857G>T]	E1953X	Heterozygous
RP349	41	7919G>A	W2640X	Homozygous

4.9 Identification of 2nd heterozygous change in RP73

Due to the identification of a heterozygous base pair change identified in family RP73, collaborators at Sanger Centre performed an Agilent array and a Nimblegen chip on members of this family. Since the mother of this family was deceased, only the fathers DNA and one affected DNA samples were analysed (See figure 1.7 for pedigree of RP73 family). As previously depicted through a restriction enzyme digest experiment, the heterozygous point mutation was not identified in either of the alleles belong to the father. It is therefore assumed that if a second change is identified through tilepath array analysis it should present in the father in a heterozygous manner i.e. only on one allele. Four clones were identified as being absent in the affected members of this family, these were mapped to the same region as the deletion of the clone in RP5 (Figure 3.12) suggesting that a second mutation has been identified in the RP73 family. This deletion was not identified in the corresponding exons of *EYS*, as full amplification was attainable in all exons in all members of this family, leading to the suggestion that this change is the second heterozygote deletion in this family. The identification of this change was imperative, as it concluded our sixth conclusive mutation in the affected families. It is interesting to note that all five families acquired varying types of mutations leading to the RP25 phenotype (Table 4.2).

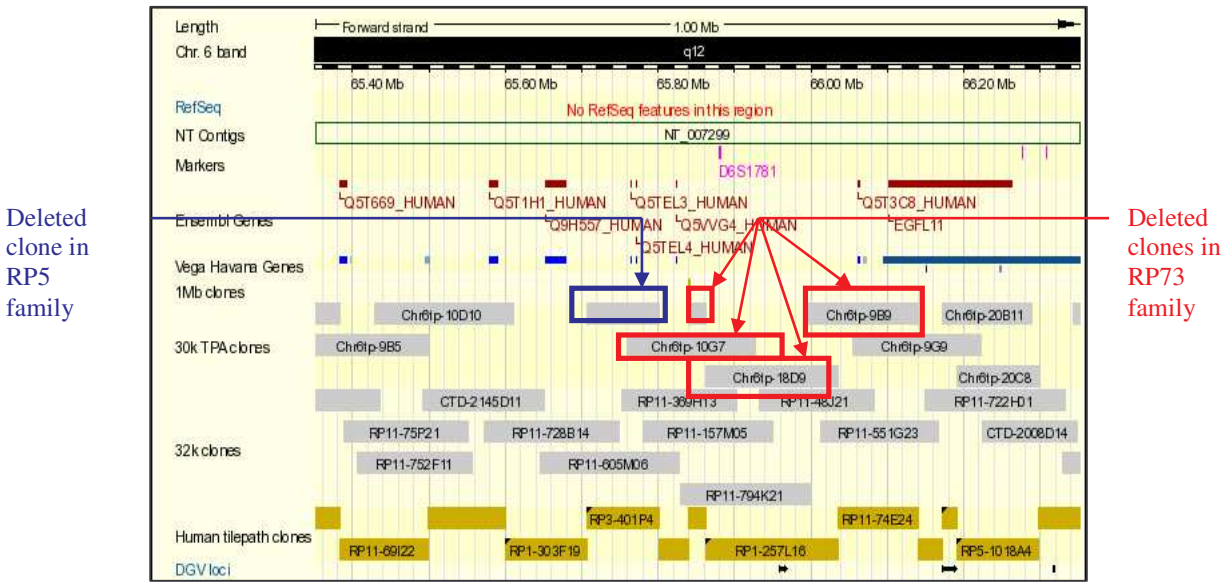


Figure 4.16 An ensembl view (Version 43: April 2007) of the clones of interest in the region. Highlighted in red are the clones identified as being deleted in the affected members of RP73, the clones include Chr6tp-10G7, Chr61p-18D9 and Chr61p-9B9. The combined size of these clones is 270 kb. The clone highlighted in blue corresponds to the clone deleted in RP5 affected members.

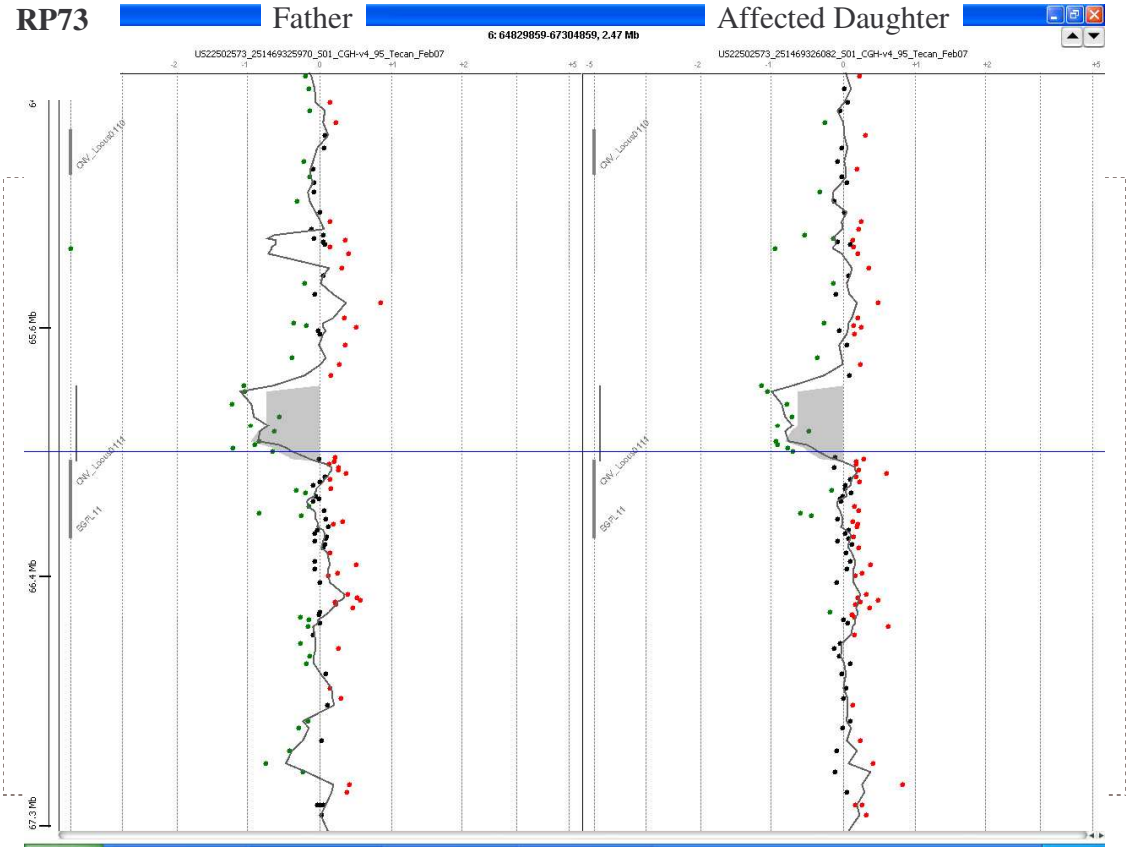


Figure 4.17 An Agilent array performed on members of the RP73 family. The above readout is of the father and the affected daughter on chromosome 6. The blue line indicates where the deleted clone was identified, the high peaks in both members demonstrates that this deletion was in both carrier father and affected daughter of this family as this is the 2nd heterozygous deletion identified. This experiment was performed by the Sanger Centre, Cambridge.

4.10 Refinement of large deletions using MLPA (Multiplex Ligation Probe Dependent) assay and Agilent arrays.

The following experiments were performed by colleagues in Seville, Spain and in the Sanger Centre, Cambridge.

Two large deletions were identified in families RP5 and RP73 respectively, whose breakpoints had yet to be ascertained. To identify these breakpoints, an MLPA assay and an Agilent array were performed, the MLPA assay was designed being performed by our Spanish colleagues and the 244k Agilent array was undertaken by the Sanger Centre, Cambridge. For the MLPA assay a set of control probes was included in each synthetic probe mix in order to provide invariable values. From these experiments, the deletion in the RP5 family was determined to span from 65581641 to 65783332 nucleotides on chromosome 6q, being 202 kb in size (Figure 4.18). The deletion in RP73 spans 270 kb, from 65816910 to 66087361 nucleotides (Figure 4.19).

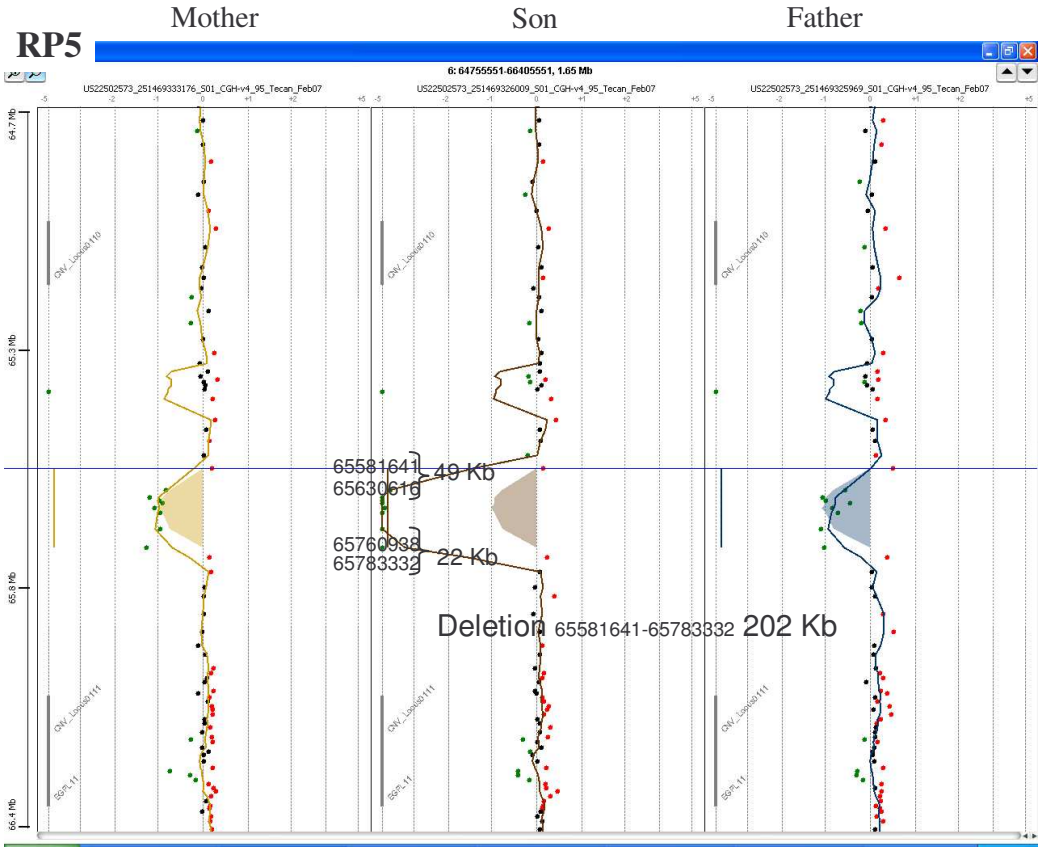


Figure 4.18 As previously depicted, an Agilent array was performed on RP5 family members. This array also narrowed down the interval to identify the deletion size and breakpoint. The deletion ran from 65581641 to 65783332, the fragment spanning 202 Kb. This was performed by the Sanger Centre, Cambridge.

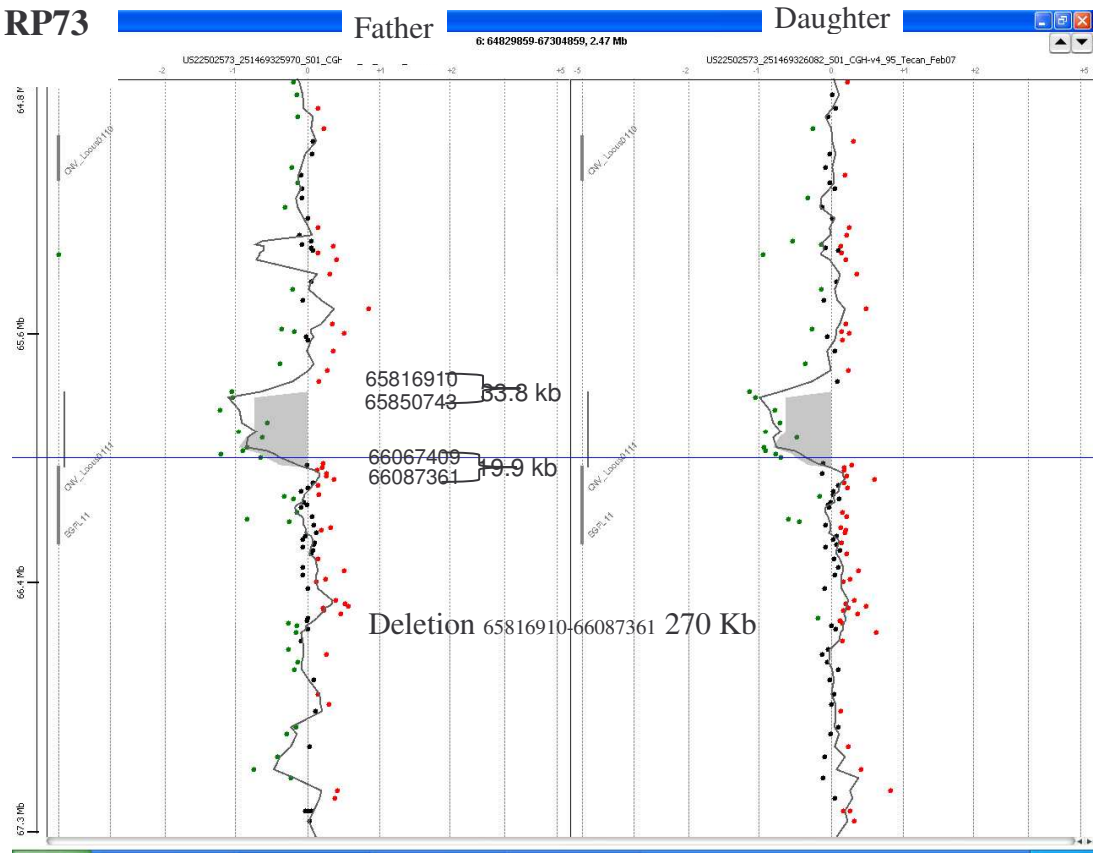


Figure 4.19 As in figure 4.18, an Agilent array was performed on family members of RP73. This array identified the exact breakpoint and size of the deletion as being 270 Kb in size, and running from 65816910 to 66087361 in chromosome 6. This was performed by the Sanger Centre, Cambridge.

4.11 Mutation screening of arRP panel

Once the full length *EYS* gene was identified, a thorough mutation screening of a mixed ethnic arRP panel was performed. Patients were British residents of a variety of ethnic origin (European Caucasian and Asian from the Indian subcontinent). All 96 patients were identified through the retinal dystrophy clinics at Moorfields Eye Hospital using standard techniques. Informed consent for genetic studies in adherence to the Declaration of Helsinki was obtained by the examining clinicians. Primers were designed to span approximately 200 base pairs aligning each exon (see appendix). In total 47 primer pairs were designed for all 43 exons, three of which are completely non-coding, a further two having a non coding region. Amplification standards for each exon can be found in appendix. Novel changes identified were compared with the Ensembl database to confirm the prevelence of the change. Once a change was identified to be novel, a control experiment was performed on 96 ethnically matched individuals with no personal or family history of retinopathy. These were selected to serve as controls. Any change which was not identified in the control population was then proposed to be a novel and a possible mutation for RP25.

Many SNPs previously annotated on Ensembl were identified across the gene, several novel SNPs were also identified in many members of the arRP panel (Table 4.1). These changes, though not identified on Ensembl, were identified through amplification of DNA from a control panel, therefore eliminating them as putative mutations. Four interesting novel changes were identified in specific arRP patients (Figure 4.21); in

exon 36, a heterozygous base pair change G>A, causing an amino acid change of arginine (R) to lysine (K). This was only identified in one patient in the panel, patient annotated number 32. Two further heterozygous changes were identified in other recessive panel members. In patient 28, a heterozygous C>G change was identified leading to an amino acid change from tyrosine (Y) to a stop codon (*). In 18 patients of the panel a G>C heterozygous change was identified causing an amino acid change of valine (V) to alanine (A). Interestingly this change was not identified through screening of the control panel. A final change was identified in exon 40, this again was a heterozygous change of C>T leading to a change of amino acid of arginine (R) to cysteine (C), identified in patient 23 of the 96 recessive panel (Table 4.3).

Since *EYS* is a recessive gene, it is necessary for both alleles to contain a mutation in order for the mutation to be considered as putative. All changes identified in this panel have been heterozygous; this does not rule out that these are putative changes; however a second change may be necessary in order for the protein to become redundant. It is suggested that the second changes in these patients may be large mutations or changes in intronic regions. A compound heterozygous change has previously been identified in RP73. Also, large mutations were identified in RP5 and RP73, therefore it would be suggested that a CGH array analysis be carried out on these patients in order to identify further mutations. For these changes to be considered as real is that further control panel screening should be performed. If time had prevailed, mutation screening of at least 200 members from a mixed ethnic population should have been performed allowing for a more accurate statistical analysis.

Table 4.3 Novel Changes and SNPs identified in arRP panel

Exon	Sequence variation	Amino Acid Change	SNP ID	Patient No	Control Panel Screened
1	c.-508 A>G c.-546-81 A> C	- _____	rs1490127 NOVEL		 12% allele frequency
4	c.-359 C>T c.- 748+52 T>C	p. T120M _____	rs12193967 NOVEL		 18% allele frequency
5	c.-862+87 T>C		rs4710522		
6	c.-863-22insTT c.911		rs34154043 rs34676630		
7	c.-1146 T>C	p.N382N	rs974110		
9	c.-1300-3C>T c.1459+103C>T		rs1936439 rs9453265		
30	c.6078-4_6078-3delTC		rs35395170		
34	c.8634+61T>G	-	ENSSNP748595		
35	c.6835-64C>T c. 6977G>A	- p.R2326Q	rs1482457 rs4710457		
36	c. 7205 G>A	p. R2402K	NOVEL	32	No allele frequency
39	c.7723+64T>A c.7666A> T c.7665 C>G c.7658G>C	_____ p.S2556C p. Y2555* p.V2553A	NOVEL NOVEL NOVEL NOVEL	In 3 patier In 5 patier 28 In 18 of 9	30% allele frequency 25% allele frequency No allele frequency No allele frequency
40	c.7796 A>G c.7810 C>T	p. P2566P p.R2604C	NOVEL NOVEL	In 3 patier 23	10% allele frequency No allele frequency

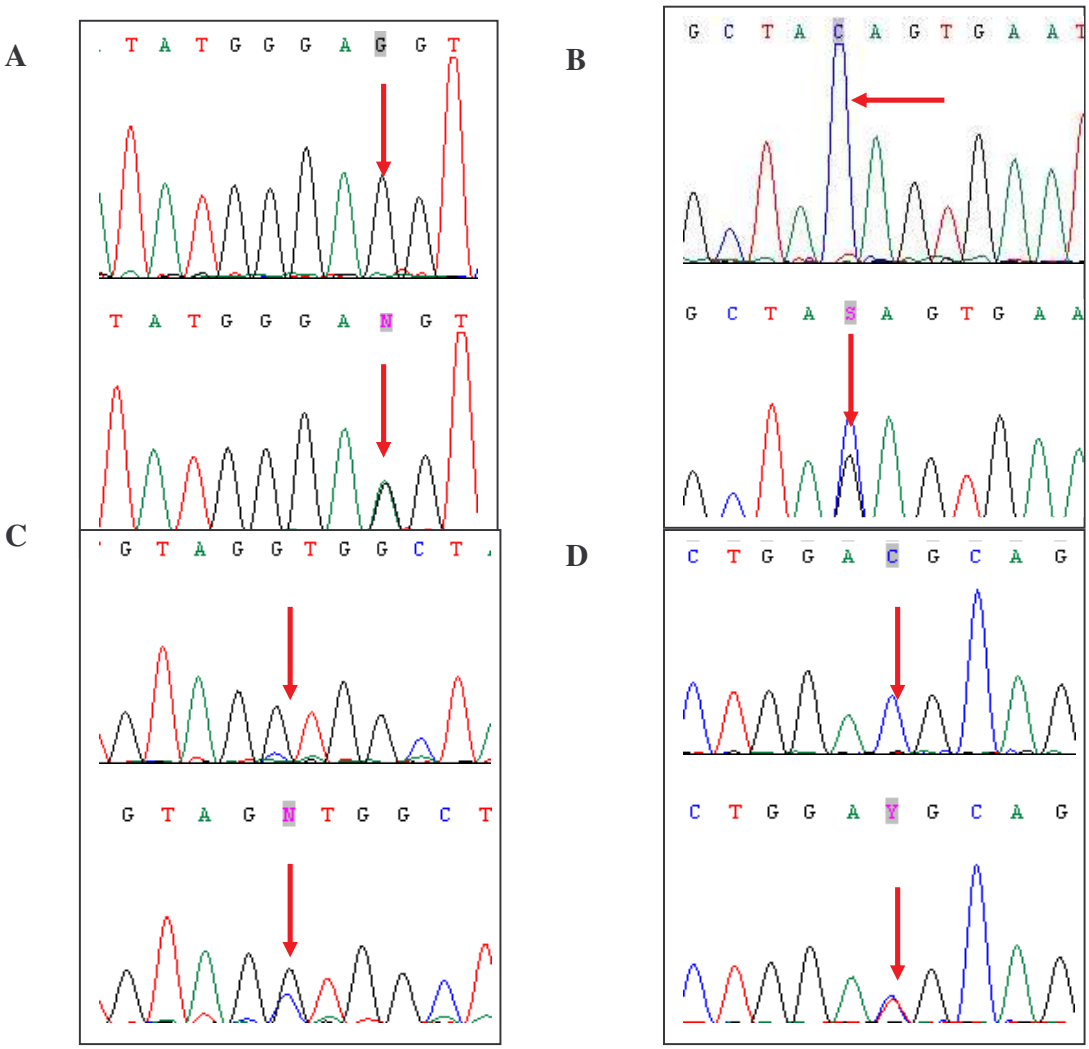


Figure 4.21 Electropherograms of novel mutations identified in individual patients of an arRP panel. A: c. 7205 G>A change identified in exon 36 in patient 32. B: c.7665 C>G change identified in exon 39 in patient 28. C: c.7658G>C identified in exon 39 in 18 members of the panel. D: c.7810 C>T in exon 40 in patient 23.

4.12 Conclusion of Chapter 4

The *EYS* transcript initially spanned 14 exons with no translation start or stop codon identified, however through further 5`and 3` RACE the gene expanded to incorporate the already characterised *EGFL-11* gene making the gene 26 exons in size. At this time, the translation start codon of the *RP25* gene was identified; however the translation stop codon was not determined. A 1,238 bp coding fragment was identified at the 3` end of exon 29, which did not alter the coding region of the gene; this novel discovery had not been previously annotated in the available human genome database. After several more rounds of 3`RACE and bioinformatics exonic predictions, the final full length transcript of the *RP25* gene was determined. On assembling all available data, *RP25* encompasses 43 exons, 30 exons belonging to 9 previously predicted genes and 13 newly reported exons, spanning from 64487835 and 66473839 MB on chromosome 6q12. The *RP25* full length transcript is retinal specific and is greater than 2.0 MB in size, thus it is the largest eye-specific gene found to date. It is also the fifth largest gene in the human genome to date.

Human *RP25* was determined to be an ortholog of the fruit fly (*D. melanogaster)* gene *eys* (Spacemaker), thus *RP25* was renamed *EYS* (SPAM). Orthologs of this gene were not identified in insects such as the honeybee, as these insects have a variation in their visual system and is not an open rhabdome visual system. An apparently intact *EYS* gene is found across the mammalian clade, including monotremes

(platypus) and marsupials (opossum). However, despite the mutations and the presumed loss of function associated with human disease, this gene has been dispensed with on at least four separate occasions in the last 100 million years of mammalian evolution, including the armadillo (*Dasypus novemcinctus*), the little brown bat (*Myotis lucifugus*) and ruminant (cattle and sheep) lineages. *Eys* has acquired many (<3) reading-frame disruptions in three rodents (mouse, rat and guinea pig) representing two of the three major rodent clades. RT-PCR confirmed the absence of *eys* in mice and rat genomes.

Mutation screening analysis of the full length *EYS* gene continued on all linked Spanish families. Three novel mutations were identified in three unrelated families. In RP73, a heterozygous null mutations was identified, later a second large deletion (270 Kb) was confirmed through CGH analysis. This was the first compound heterozygous change revealed thus far in *EYS*. Two further unique mutations were identified in two unrelated families respectively. In family RP328, a homozygous base pair deletion was identified causing a frameshift in the gene. In family RP349, a homozygous null mutation was identified. The breakpoints of the deletion in the RP5 family were also determined to be 202 Kb through MLPA and CGH analysis. In total, five unrelated Spanish arRP families with six unique mutations, one being a compound heterozygote, all leading to the RP25 phenotype.

Once the full length gene was identified, arRP panel mutation screening of *EYS* began. From this, four novel heterozygous mutations were identified. These could be

considered as putative changes, however for these changes to be considered as disease causing for recessive RP, a second mutation would need to be identified, therefore considering these changes as compound heterozygous.

CHAPTER 5

Immunolocalisation of SPAM

5.1 Production of SPAM antibody

Once the full length gene and protein structure of SPAM were ascertained an antibody was designed against the protein. Antibodies demonstrate both the presence and the sub cellular localisation of a protein through the antigenic epitopes present in the polypeptide. The antibody was selected through the most suitable predictions offered by Eurogentec S.A (Belgium). Antigen selection is usually based on hydrophilicity, amino acid composition, and certain post-translational modifications (Pow *et al.,* 2008).

An antigen against the amino-terminus of the human protein was selected as being the most suitable; this was done with the expertise of Professor Mike Cheetham, UCL. A peptide sequence for the N-terminal sequence of the protein was selected due to the amino- and carboxyl-termini being solvent-exposed sequences and mobile in crystallographic structures of proteins. Rabbit polyclonal antiserum was raised against one peptide derived from the predicted amino acid sequence by Eurogentec. The

peptide (human N-terminal residues 61-75, GVNTKIDTSGNQAVP) was conjugated to hemocyanin through an N-terminal cysteine residue before immunization. The length of the chosen peptide is an important factor, as peptides with residues of nine or less may not be effective antigens. Similarly, peptides longer than 16 amino acids may contain several epitopes; the peptide selected in this instance is assembled of 15 amino acids. A final bleed from the immunised rabbit was affinity purified against a peptide made from Eurogentec (Double XP program), and the affinity purified serum was used in all subsequent experiments.

5.2 Immunocytochemistry with SPAM

5.2.1 SPAM localisation in Y79 cells

The existence of *EYS* in human retinoblastoma cell line Y79 was confirmed in chapter 4. To validate this result and to examine the specificity and quality of the SPAM antibody, ICC was attempted using these cells. Y79 cells acquire morphological and cytochemical properties typical of human photoreceptors (Tsokos *et al.,* 1986). Because these cells are a suspension cell line, additional steps during the ICC protocol were required to obtain positive results (Section 2.9).

A gradient of SPAM antibody dilutions were attempted, in this instance, the optimum condition for anti-SPAM was a 1:200 dilution with a Cy3 (red) conjugated AffiniPure goat anti rabbit secondary antibody (1:400, JacksonImmunoResearch lab). Negative

control experiments were performed in parallel with cells being stained with the Cy3 secondary antibody alone with results revealing a positive punctate speckled staining of anti-SPAM in the Y79 cells (Figure 5.1A). Minimal background staining in the negative control experiment (Figure 5.1B) contributes to the validity of the positive anti-SPAM staining in these cells.

To further confirm the positive staining of SPAM in Y79 cells, a co-localisation experiment was performed with alpha tubulin. Alpha tubulin is an intercellular filament structure which makes up cells' cytoskeleton and exists throughout the cell. In this instance, it is used as a marker to identify how abundant the anti-SPAM signal is. Y79 cells were stained with anti-SPAM (1:200) and anti-alpha tubulin (1:200, personal communication with Professor Matter, Institute of Ophthalmology, UCL). SPAM staining (red) revealed an abundant punctuates speckled signal throughout the cytoplasm of these cells, with alpha tubulin (green) acting as a marker around the cells. A yellow signal is indicative of a possible overlap of staining of anti-SPAM and anti-alpha tubulin (Figure 5.2). All ICC experiments were visualized using a Zeiss LSM510 laser scanning confocal microscope. These results are preliminary and do not give comprehensive detail into the precise localisation of SPAM in the cell. One conclusion is that there is a high level of SPAM expression in the cytoplasm of Y79 cells. This is consistent with the cDNA expression profile which illustrates that *EYS* mRNA (SPAM being the protein) is present in abundance in this retinoblastoma cell line (Figure 4.4).

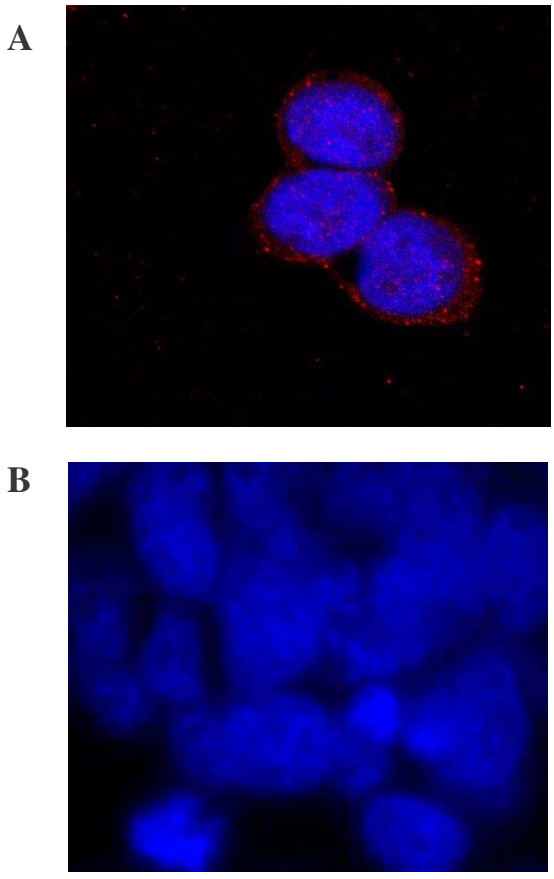


Figure 5.1 ICC staining of Y79 cells with anti-SPAM alone. **A** anti-SPAM staining (red) highlights a prominent speckled staining throughout the cytoplasm of the cell. **B** A control experiment performed using Cy3 secondary antibody alone, only highlighting DAPI (4', 6-diamidino-2-phenylindole) staining of the cell nuclei.

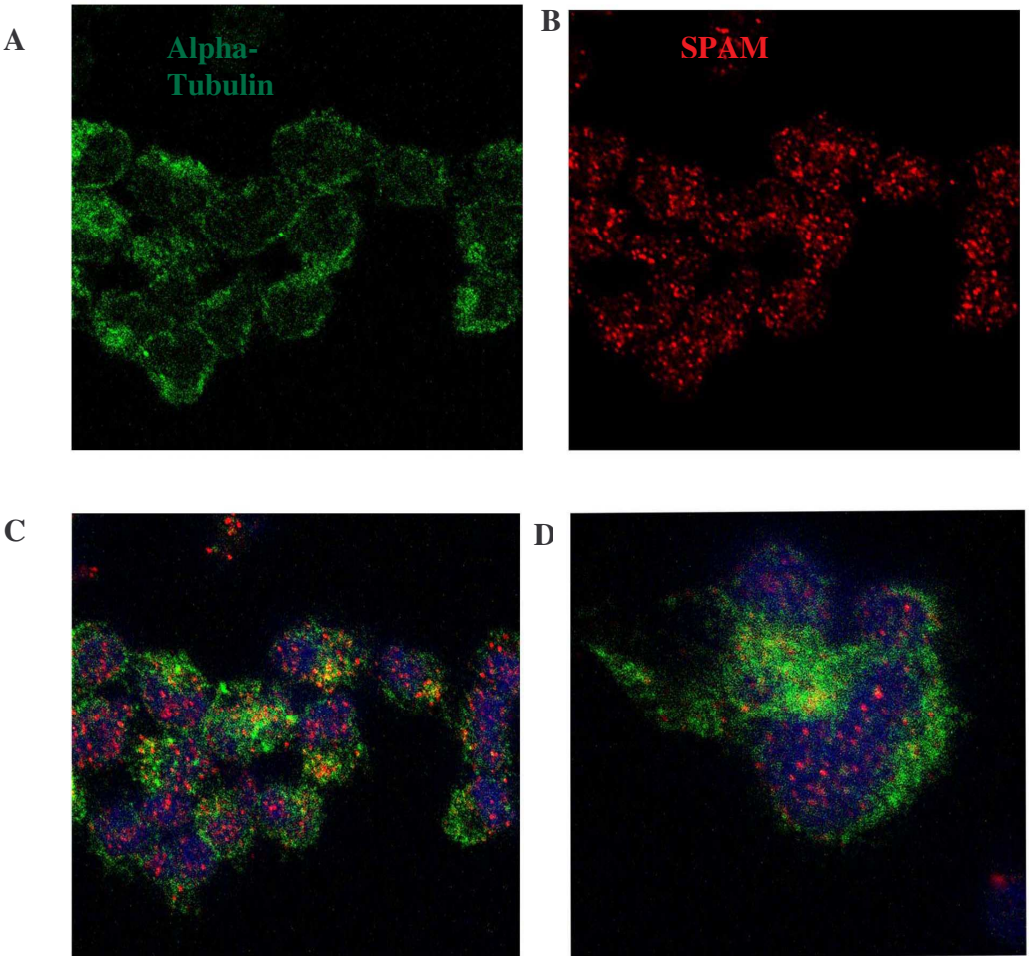


Figure 5.2 Immunofluorescent staining of Y79 cells with anti-alpha-tubulin (green) and anti-SPAM (red). **A** ICC staining for anti-Alpha tubulin (FITC). **B** A signal for anti-SPAM (Cy3), highlighting a speckled staining prominent in the cytoplasm. **C** Overlay of alpha-tubulin (green), SPAM (red) and DAPI (blue) staining. A yellow signal highlights a possible overlap of signal from the SPAM and alpha tubulin antibodies. **D** Enlarged view of a single Y79 cell stained with SPAM and alpha tubulin antibodies.

5.3 Immunohistochemistry (IHC)

5.3.1 IHC on mouse retina tissue

The mouse is a key model organism for studying diseases such as retinitis pigmentosa; it has a well characterised genome with most genes identified as having human orthologs (Huang *et al.,* 2004). As described in chapter 4, SPAM is proposed to be dispensed with of in the mouse genome. To validate this assumption further, IHC was performed on mouse retina sections.

When the epitope raised for SPAM was compared against the mouse genome, no significant similarity was detected; this is proven through IHC experiments on mouse retina tissue. Co-localisation staining experiments were performed using well characterised antibodies such as 1D4, anti-TOPORS (Abnova, UK), anti-RP1 (kindly provided by Professor Mike Cheetham, UCL) and PNA (details on all antibodies can be found in Section 5.7). All control experiments were performed in parallel. Anti-SPAM was stained using a goat anti rabbit FITC secondary antibody. Numerous attempts using a gradient of SPAM antibody concentrations were attempted. A prominent fluorescent signal was achieved using the control antibodies (Figure 5.3); however when performing IHC with the SPAM antibody, no significant signal was achieved. All IHC experiments were visualized using a Zeiss LSM510 laser scanning confocal microscope. This result further validates the proposal that SPAM is absent from the

mouse genome. However, it may be argued that the lack of signal detection could be due to the low similarity between the human SPAM sequence and its corresponding ortholog. In this particular case, the lack of signal detection is due to the mouse SPAM ortholog being a pseudogene.

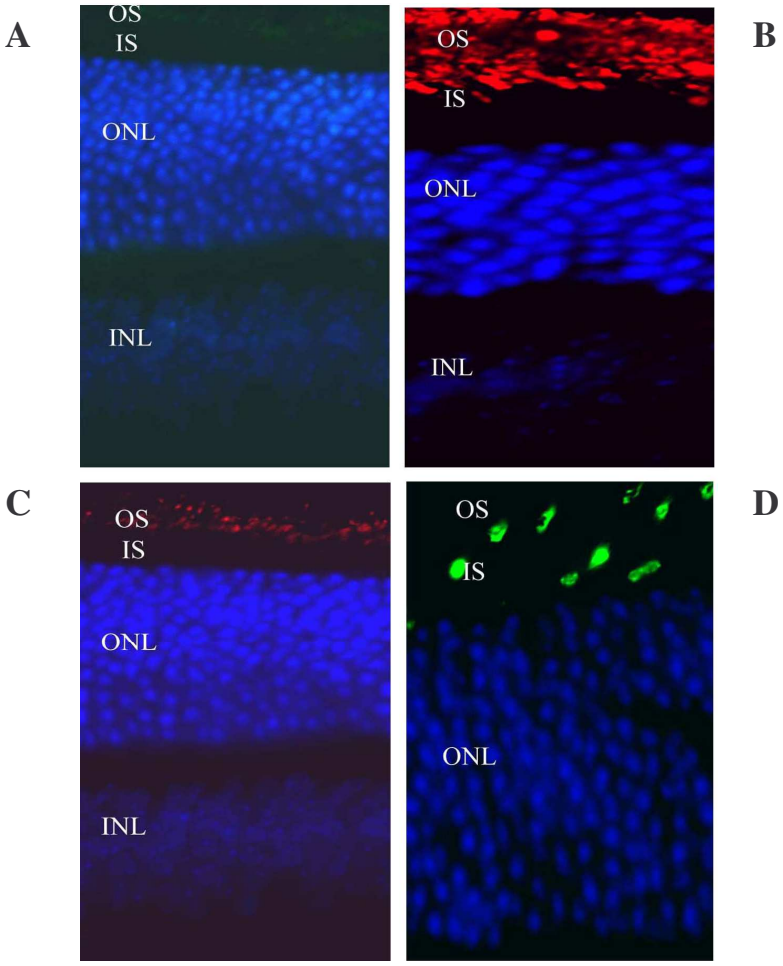


Figure 5.3 Mouse retina IHC staining with anti-SPAM, RP1, 1D4 and PNA antibodies. **A** anti-SPAM staining alone (green). **B** anti-SPAM (green) co stained with 1D4 (red). **C** anti-SPAM (green) is co stained with anti-RP1 (red). **D** anti-SPAM (red) co stained with PNA (green). There is an obvious lack of signal from the SPAM antibody in all experiments. OS- outer segment; IS- inner segment; ONL- outer nuclear layer; INL-inner nuclear layer.

5.3.2 IHC on porcine retina tissue

Based on the prediction that SPAM is absent from the mouse genome, an alternative source of retinal tissue had to be chosen. Selection was based on BLAST similarity searches across several candidate genomes. BLAST was performed using the epitope sequence of the SPAM antibody as a target sequence and BLAST searches were carried out on pig (*Sus scrofa*), chicken (*Gallus gallus*) and rabbit (*Oryctolagus cuniculus*). Of these, pig was selected as the sequence identity of the antibody epitope. The gene LOC100153323, similar to EGF-like-domain, multiple 11 (*Sus scrofa*) which is the ortholog of SPAM in the pig, was identified as having 64% similarity for this peptide sequence (Figure 5.4). One of the major parameters which make an antibody successful or not is its specificity, even though the ortholog in this instance was 64% it seemed to be adequate to detect a signal for SPAM. Initial attempts to detect a signal failed due to factors such as autofluorescence which was redeemed by using an adequate pre-treatment protocol (Section 2.8).

Control experiments are a critical aspect when characterising any antibody. Possible cross reactivity is eliminated by performing double labelling experiments using one primary antibody against the secondary of the other antibody. Another fundamental control experiment is conducted using the antibody in combination with the competing

peptide, which is also known as immunoneutralisation. The purpose of this experiment is to ensure the signal identified is authentic and can be blocked when combined with the peptide.

Conclusive results can be assumed from these preliminary porcine retina IHC experiments. Figure 5.5 highlights a significant staining in the photoreceptors of porcine retina tissue using anti-SPAM with a FITC secondary antibody (green). Successful staining was achieved using anti-SPAM in a 1:200 dilution. The lack of staining elsewhere in the retina leads to the initial postulation that SPAM is primarily located in the photoreceptors. All controls were successfully completed, including immunoneutralisation (Figure 5.6 B), where anti-SPAM staining appears as a faint speckled signal in the tips of the photoreceptors.

```
>    ref|XP_001924954.1| UG PREDICTED: similar to EGF-like-domain,
multiple 11 [Sus scrofa]
Length=294

 GENE ID: 100153323 LOC100153323 | similar to EGF-like-domain,
multiple 11
[Sus scrofa]

 Score = 29.5 bits (62), Expect = 0.066
 Identities = 9/14 (64%), Positives = 10/14 (71%), Gaps = 0/14 (0%)

Query 2  VNTKIDTSGNQAVP 15
         VN KI+TS N  VP
Sbjct 63 VNAKINTSENHVVP 76
```

Query- Sus scrofa.

Sbjct- SPAM peptide.

Figure 5.4 BLAST results of the peptide sequence for the SPAM amino-terminal antibody against the pig genome. A 64% identity between the porcine ortholgoue and human peptide was observed.

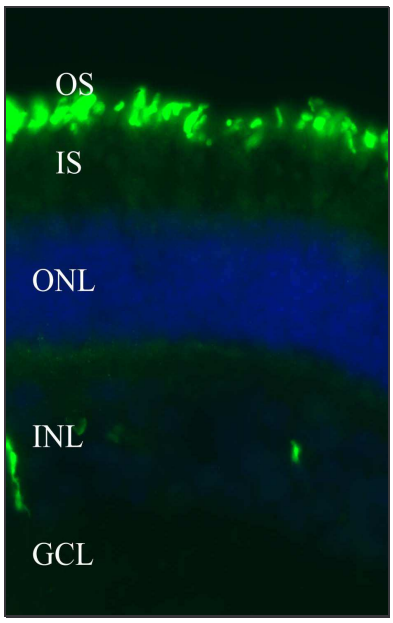


Figure 5.5 IHC staining on a porcine retinal section highlighting anti-SPAM (green) staining using FITC goat anti-rabbit secondary antibody. A strong signal is identified in the photoreceptor cells. DAPI staining in blue highlights the nuclear layers of the retina. OS-outer segment; IS-inner segment; ONL-outer nuclear layer; INL-inner nuclear layer; GCL-ganglion cell layer.

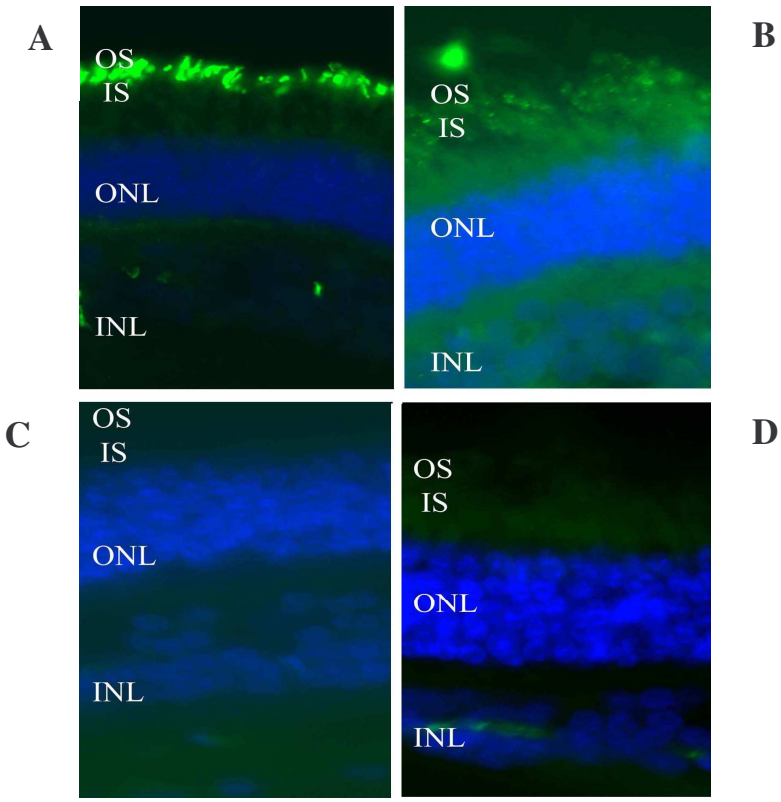


Figure 5.6 Control IHC experiments on porcine retinal sections. **A** anti-SPAM staining, highlighting a strong signal for anti-SPAM. **B** SPAM peptide versus SPAM antibody, demonstrating a speckled appearance, highlighting the blocking of the antibody by the competing peptide. **C** Primary SPAM antibody without a secondary antibody, no signal detected. **D** FITC goat anti-rabbit secondary antibody alone, without any primary antibody therefore no signal was detected. **C** and **D** highlight that there is no background or secondary antibody staining. DAPI staining highlighting nuclear cell layers.

5.4 Co-localisation of SPAM with other proteins

5.4.1 SPAM and Rhodopsin

Immunofluorescence staining experiments are also successfully used for labelling multiple antigens in the same preparation. This method is particularly successful for the co-localisation of antigens in the same site of the tissue section. An initial co-localisation staining was performed using an anti-rhodopsin antibody, 1D4. Rhodopsin (*RHO* or *RH1*) is a well characterised gene and is a major cause of adRP and some arRP cases. It is a pigment of the retina that is responsible for both the formation of the photoreceptors and is also involved in the perception of light, detecting the levels of illumination. RHO is identified to be tightly packed in the outer segments of rod photoreceptor cells (Kisselev, 2005) where it can maximise its photon absorption. 1D4 is a monoclonal antibody that specifically detects the highly conserved C-terminus of rhodopsin molecules from a wide variety of mammalian (and many non-mammalian) rod cells. 1D4 was raised in mouse, while anti-SPAM was raised in rabbit. All control experiments were performed in parallel.

The result of this experiment is extremely significant as it appears that SPAM and RHO completely co-localise producing a strong yellow staining when double labeled, with RHO staining being more prominent on the rim of the outer segments of the photoreceptors (Figure 5.7). From this result it is possible to postulate that SPAM localises specifically to retinal photoreceptors, due to a signal not being detected in any other cell layer.

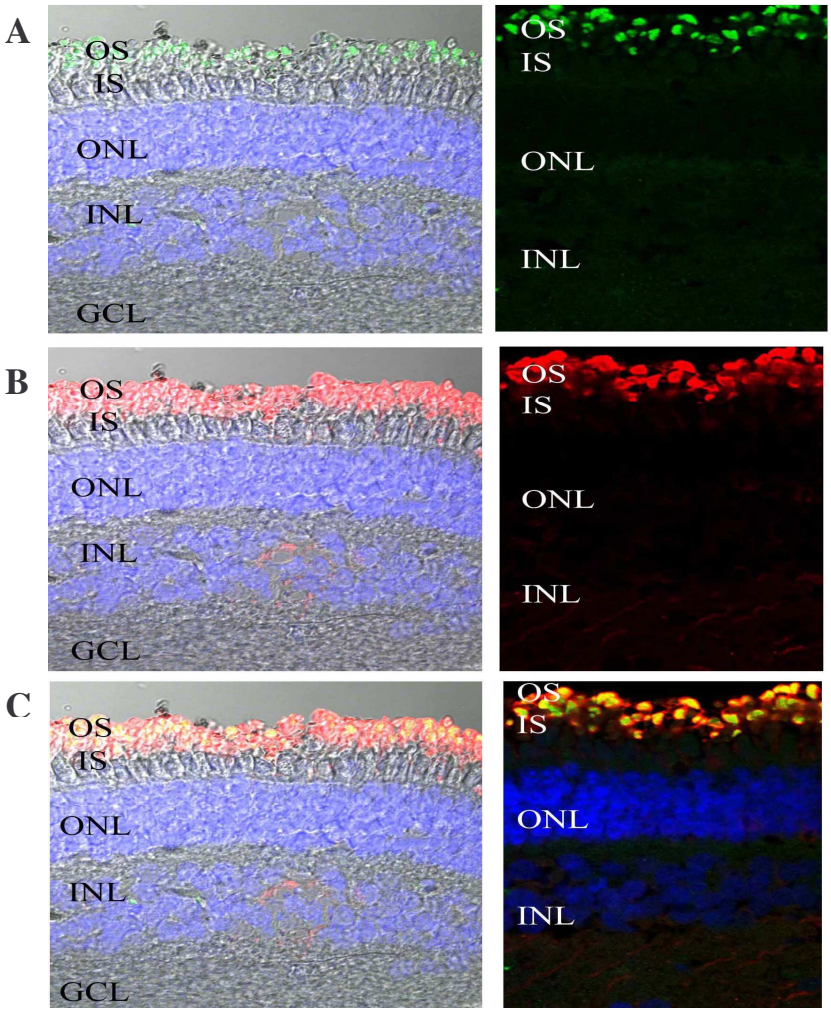


Figure 5.7 IHC co-localisation of SPAM and RHO antibodies in porcine retina. **A** IHC staining with anti-SPAM (green). **B** Staining of 1D4 (anti-RHO) (red). **C** Co-localisation staining of anti-SPAM and 1D4, with the nuclear layer. The yellow colour depicts the overlap of signal of anti-SPAM and 1D4 in the outer segments of photoreceptors.

5.3.2 Localisation of SPAM with PNA and WGA

To further speculate the precise localisation of SPAM in porcine retina tissue, co-staining with labeled lectins PNA (Peanut agglutinin) and WGA (Wheatgerm agglutinin) streptavidin conjugates (ZyMAX Grade 1:100, Zymed laboratories, San Francisco, CA) was performed. WGA is a lectin from wheat germ that binds to N acetylglucosaminyl (Naeem *et al.,* 2007) and PNA is a lectin that specifically binds to Gal-beta (1-3)-GalNAc carbohydrates (Krull 2000). Both PNA and WGA are specific lectins that bind in the insoluble interphotoreceptor matrix (IPM) of the human retina. PNA selectively binds to the cone matrix domains whereas WGA binds to the matrix domains surrounding rods (Tien *et al.,* 1992). Antibodies utilised for both lectins were performed with a biotinylated FITC secondary antibody to ensure strong specific binding of the streptavidin. IHC staining of porcine retina with anti-SPAM and PNA resulted in a lack of co-localisation as both the antibody and marker exhibit independent signals with no overlap of staining (Figure 5.8). This leads to the hypothesis that SPAM is absent from cone photoreceptors in porcine, alternatively SPAM may exist in cones but at a lower expression level in comparison to the strong clearly visible staining in rod photoreceptors. Double IHC labelling with WGA reveals a strong yellow orange staining concluding the foremost existence of SPAM is in rod photoreceptor cells (Figure 5.9). This antibody staining result conclusively exhibits the presence of SPAM in rod photoreceptor cells.

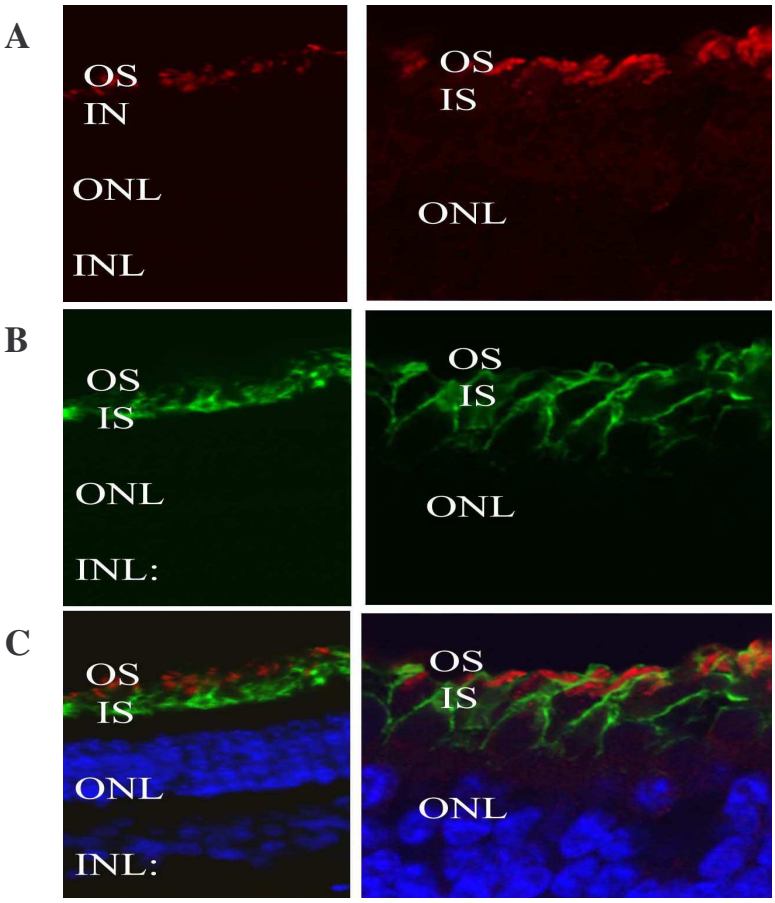


Figure 5.8 IHC co-staining of SPAM and PNA (stains the rims of cone photoreceptors). **A** Anti-SPAM (red). **B** PNA (green) staining. **C** Overlay of both antibodies with DAPI nuclear staining. There is no colocalisation of SPAM and PNA. The right hand panels depict an enlarged view of PNA and anti SPAM staining.

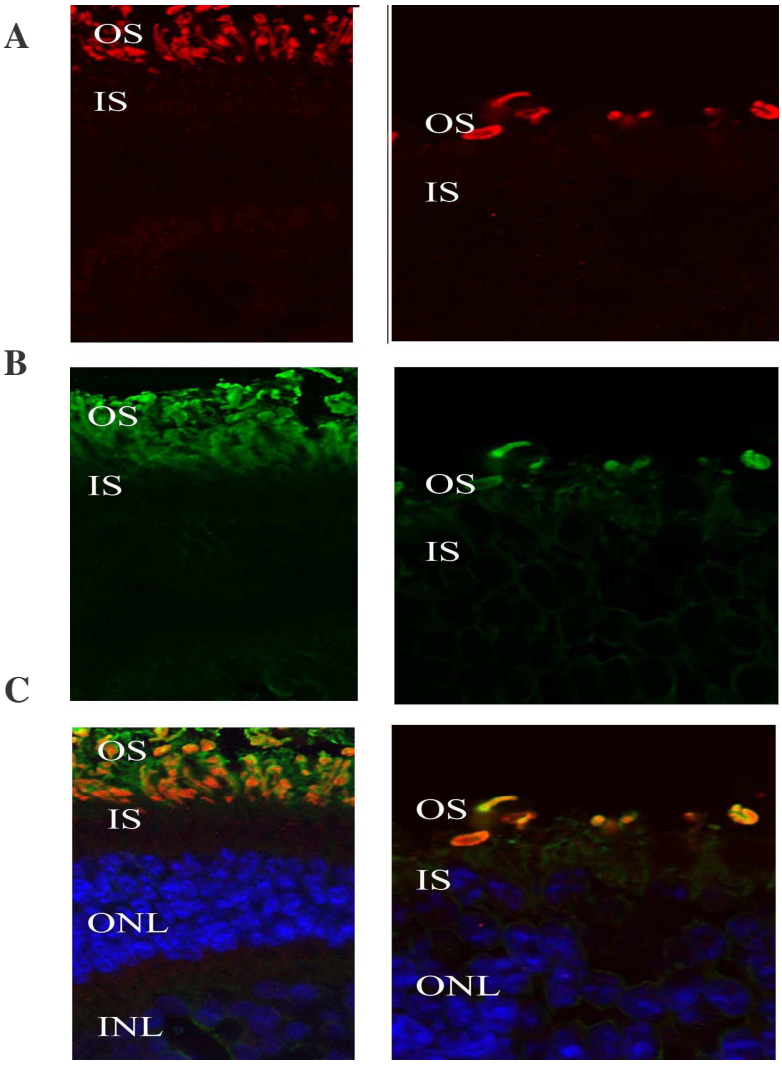


Figure 5.9 IHC staining with anti-SPAM and WGA antibodies. **A** IHC staining of anti-SPAM (red). **B** WGA (green) staining. **C** Overlay of SPAM and WGA staining with DAPI. Enlarged view in the right hand columns of single photoreceptors stained with WGA and SPAM. The merged image of both antibodies (yellow) confirms there is some co localisation of SPAM and WGA.

5.4.3 IHC of SPAM, TOPORS and RP1 antibodies

The localisation of SPAM to photoreceptors can now be concluded, in particular to the rod photoreceptors of porcine retina. Further co-staining IHC experiments were performed to refine the precise localisation of SPAM in photoreceptor cells. Antibodies to TOPORS and RP1 were selected to define this precise localisation. *TOPORS* is a causative gene for adRP which was cloned in our lab (Chakarova *et al.,* 2007) and localises to the connecting cilium of photoreceptors. *RP1*, when a mutated result in adRP and like TOPORS is also present in the connecting cilium (Liu *et al.,* 2002). The connecting cilium is a non-motile structure that links the inner and outer segments of photoreceptors. Through performing these experiments it would be possible to deduce whether SPAM localises to the inner segments, outer segments, and the connecting cilium or in all three layer, and to observe whether there is any co-localisation between SPAM and these connecting ciliary antibodies.

Both experiments were performed in parallel and produced similar results. Co-staining of SPAM and TOPORS antibodies revealed an obvious distinction of where both proteins are localised (Figure 5.10) with no overlap of the two antibody signals. Co-staining with anti-RP1 revealed a similar result with no overlap of the antibody signals (Figure 5.11). From these two experiments it is evident that SPAM specifically localises to the outer segments of photoreceptors in porcine retina.

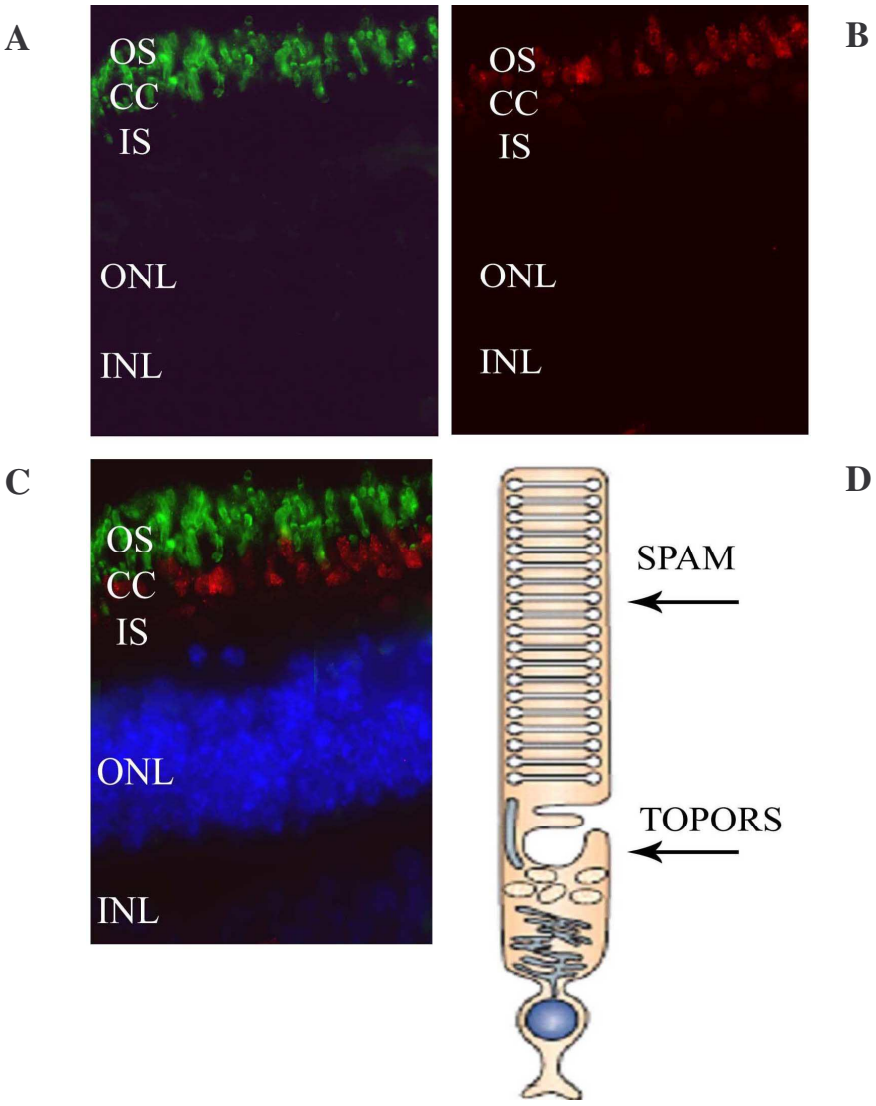


Figure 5.10 IHC staining of anti-SPAM and anti-TOPORS. **A** Anti-SPAM staining in green highlighting the outer segments of photoreceptors. **B** Anti-TOPORS staining in red highlighting a signal in the connecting cilium of photoreceptors. **C** Overlay of both staining, with no co-localisation of SPAM and TOPORS being observed. **D** A schematic view of a rod photoreceptor with arrows denoting where SPAM and TOPORS are localised.

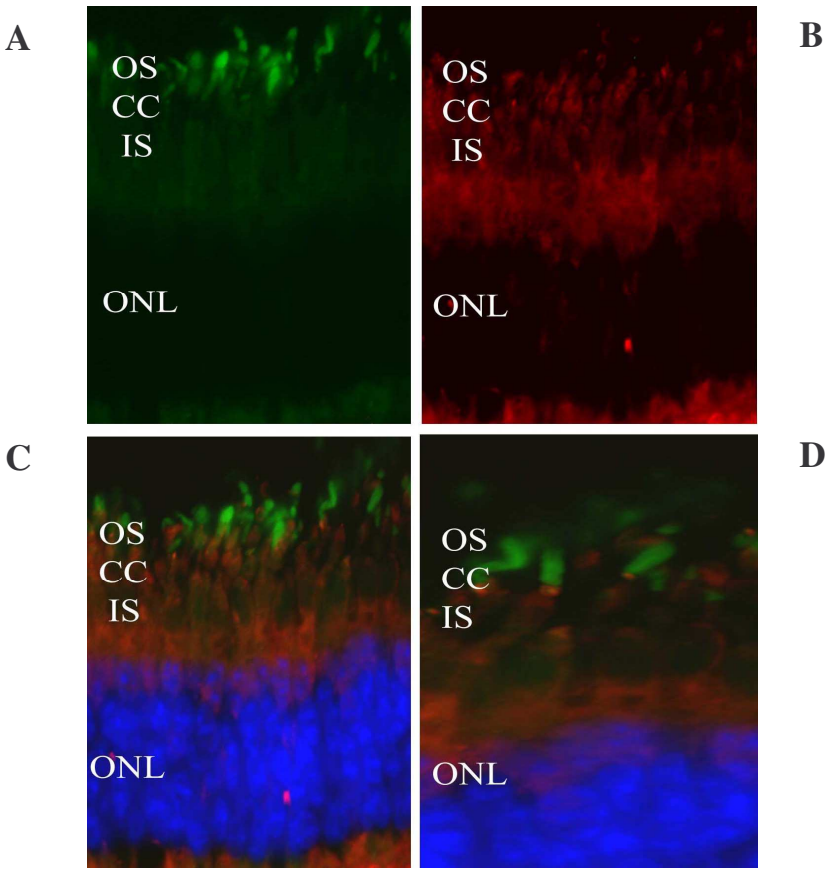


Figure 5.11 IHC co-localisation staining of anti-RP1 and anti-SPAM. **A** Anti-SPAM staining alone (green). **B** Anti-RP1 staining alone (red). **C** Overlay of both RP1 and SPAM staining with DAPI staining of the nuclear layers. **D** Enlarged view of anti-RP1 (red) staining in the connecting cilium and anti-SPAM (green) staining in the outer segments of photoreceptors.

5.5 Electron Microscopy of SPAM

Electron microscopy (EM) was performed to confirm the previous results identified in the IHC experiments and to reaffirm the existence of the signal detected for anti-SPAM in the outer segments of photoreceptors. Porcine retinal sections were cryofixed and sectioned to slices of approx 60 nm thick. Anti-SPAM was labeled with gold particles and sections were scanned on a transmission electron microscope. High magnification images of the rod outer segments (ROS) of porcine retina depict a signal for anti-SPAM, the signal displayed as black speckles (gold particles) reaffirms the expression of SPAM in ROS (Figure 5.12). In an attempt to further refine the precise location of SPAM expression in ROS a co staining EM was performed using the 1D4 antibody. A control experiment using the secondary antibodies alone were performed in parallel (Figure 5.14 A).

1D4 is specific towards the C terminus of rhodopsin (Molday and MacKenzie., 1983) and localises to the cytoplasmic surface of native photoreceptor discs (Liang *et al.,* 2003). Double gold particle staining was performed using antibodies specific to SPAM and RHO. Anti-SPAM was labeled with gold particles of 10 nm in size; 1D4 was labeled with gold particles of 5 nm. Figure 5.13 demonstrates labelling of both antibodies in the outer segment of porcine rod photoreceptors. 1D4 staining has previously been observed at the periphery of a burst photoreceptor disc (Liang *et al.,* 2003). Double staining of a burst disc demonstrates anti-SPAM and 1D4 staining at the

periphery of the disc (Figure 5.14). This labelling may also be due to the design of the antibody, if the antibody was designed to the C terminus of SPAM then this expression may have been across the full surface of the disc membranes. An example depicting this occurrence is from two antibodies designed for RHO; these are 1D4, designed against the C terminus and 4D2, designed against the N terminus (Liang *et al.,* 2003). As previously depicted, 1D4 localises to the periphery of the disc, however 4D2 has even distribution across the surface of the disc. Though it can be concluded that SPAM in this instance, localises to the periphery of rod discs it does not exclude localisation of the protein across the surface of the disc membrane.

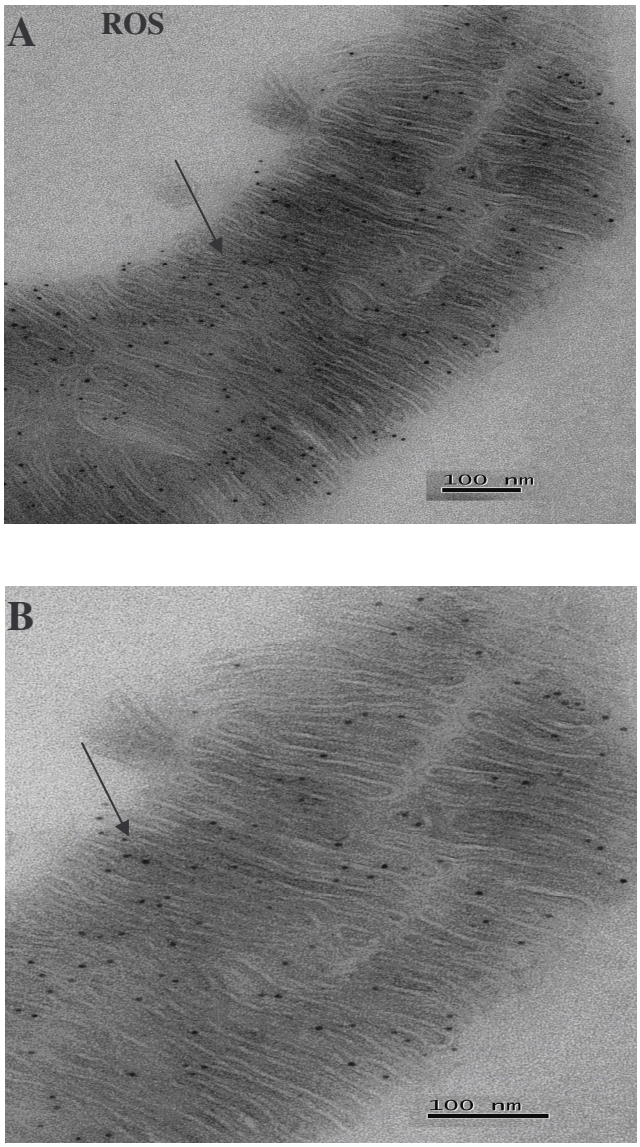


Figure 5.12 EM micrographs of outer segments of porcine photoreceptors. Retinal sections were labeled with the SPAM antibody and secondary donkey anti-rabbit antibody conjugated with 10-nm gold particles. **A** An individual rod outer segment (ROS), arrows denotes the gold labelling of the SPAM antibody localisation. **B** Enlarged view of ROS captured in A.

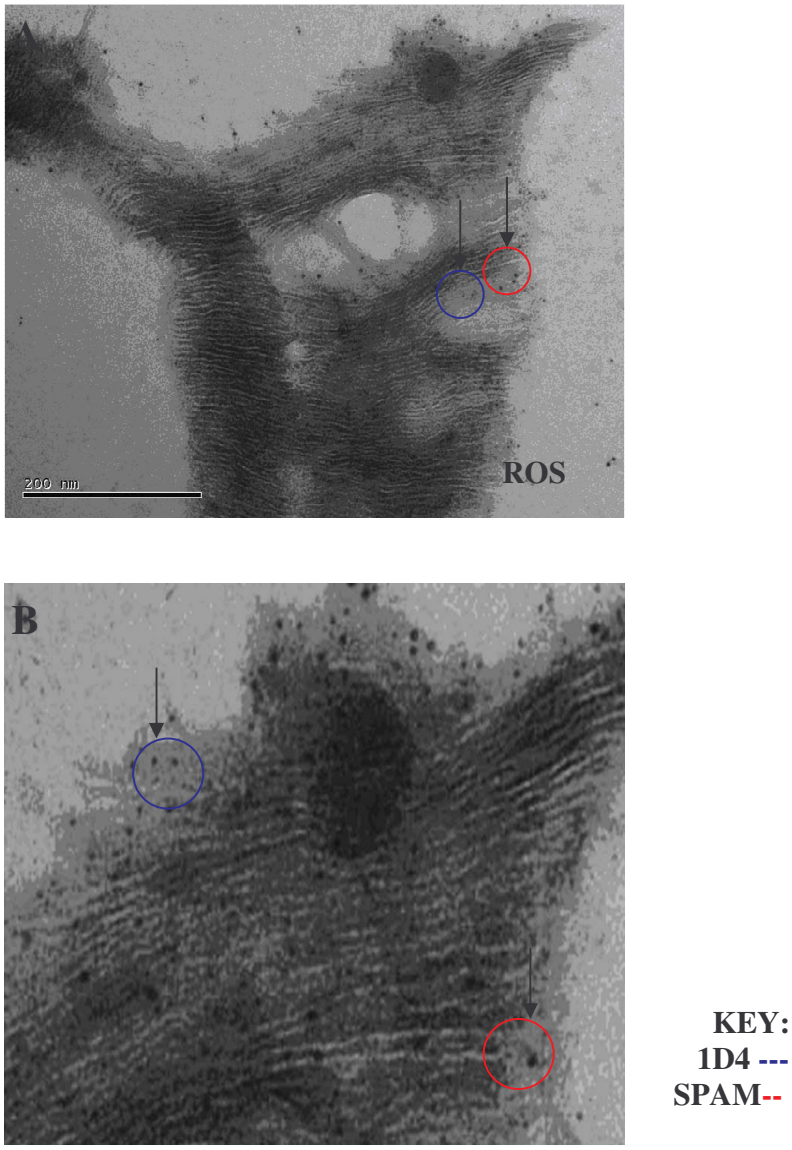
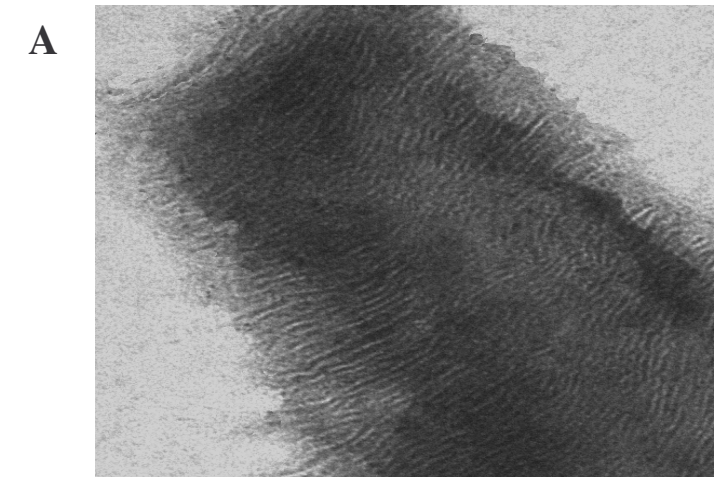


Figure 5.13 EM micrographs of SPAM and 1D4 in the retina of porcine. **A** Retinal ROS double labeled with SPAM and 1D4 antibodies, conjugated with 10-nm and 5-nm gold particles respectively. Blue circle highlights 1D4 gold particle staining, red circle highlighting SPAM gold particle staining. **B** Enlarged view of EM micrograph observed in A.



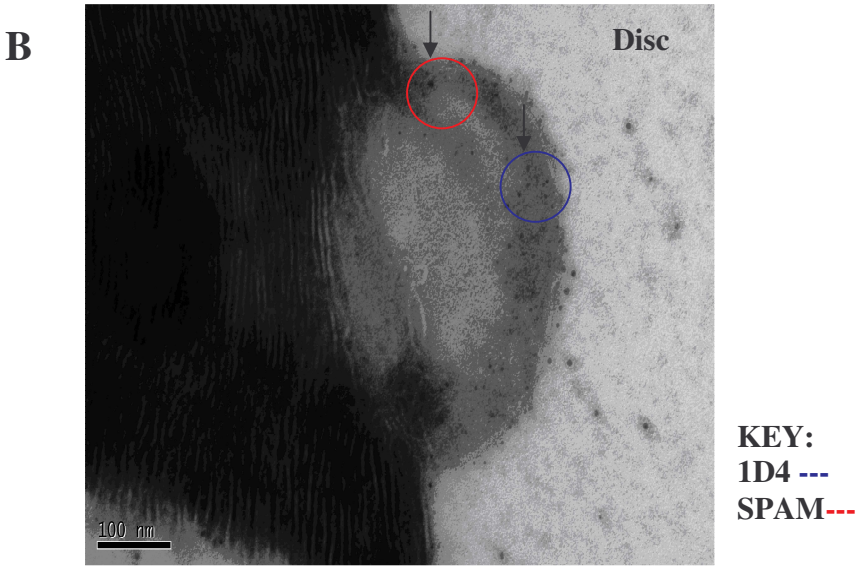


Figure 5.14 A Control experiment of ROS with secondary gold particle staining only. **B** Membrane from burst disc exposing the extracellular surface. Section was incubated with 1D4 and SPAM. Gold particles for SPAM and 1D4 are observed at the periphery of the disc.

5.6 Conclusion of Chapter 5

An antibody was designed and raised against the amino-terminus of the SPAM protein. ICC was performed using this antibody on retinoblastoma cells (Y79s), which had previously been used in a cDNA expression analysis of SPAM. Prominent expression of SPAM was determined in these cells with a strong punctuate signal being detected in the cytoplasm. This result does not exclude nuclear signal staining in these or other cells, but is a preliminary result reinforcing the expression data from chapter 4.

To reiterate the prediction that SPAM is absent in the mouse genome, IHC staining was performed on mouse retinal tissue with the conclusive result that detection of SPAM is unattainable through using this antibody. Bioinformatics predictions selected the porcine genome as having an ortholog to SPAM with a high similarity (64%) to the peptide to which the antigen was raised against. Porcine retinal sections were prepared and several IHC experiments performed. The initial indication of a prominent signal was detected using anti-SPAM at a 1:200 dilution. Since this peptide worked at 64% identity to human peptide, it highlights the fact that peptide identity need not be 100% and that the amino acids that were identical were enough to detect a signal. Co-localisation studies were then performed with a RHO antibody (1D4). Complete co-localisation of the two antibodies revealed that SPAM is exclusively localised to the photoreceptors of porcine retina. Through performing IHC with bioatynulated markers PNA and WGA, a further conclusion was drawn; that SPAM is primarily expressed in rod photoreceptors in comparison to cone photoreceptors. SPAM expression may still exist in cone photoreceptors however the foremost level of

expression exists in rod photoreceptors. Performing co-localisation staining experiments with TOPORS and RP1 antibodies further refined the expression of SPAM. From these experiments, it can now be concluded that SPAM is exclusively expressed in the outer segments of the photoreceptor cells.

The expression of SPAM in the outer segments of ROS (rod outer segments) was further affirmed through EM studies with a gold staining labelling reaction, which depicts that SPAM is expressed throughout the outer segment of rod photoreceptors at a prominent level. Double gold particle staining with 1D4 reveals the expression of SPAM is to the surface of photoreceptor discs suggesting it is localised to the periphery of disc membranes, however this result could be biased based on the design of the antibody and possible SPAM expression may be detected across the full surface of disc membranes.

Despite the identification of many genes for recessive RP, the combined frequency of mutations from all identified genes accounts for only a third of arRP patients, with individual genes contributing to 1-5% of disease (Abd-El Aziz et al., 2008) this implies that causes of arRP are as yet unidentified in the vast majority of patients.

CHAPTER 6

Discussion

Recessive RP consists of a group of disorders that include juvenile or early-onset forms of RP and overlap clinically with Leber's congenital amaurosis (LCA). Despite the identification of many genes for recessive RP (section 1.12), the combined frequency of mutations from all identified genes accounts for only a third of arRP patients, with individual genes contributing to 1-5% of disease (Abd-El Aziz *et al.,* 2008) This implies that the causative gene for the majority of arRP patients has yet to be discovered.

Genes that cause arRP encode proteins that have a range of functions in various pathways in the retina. These include being involved in the phototransduction cascade; *PDE6A* (Huang *et al.,* 1995), *RHO* (Dryja *et al.,* 1990) and *RGR* (Chen *et al*., 1996) or have a role in Vitamin A metabolism; *ABCA4* (Allikmets *et al.,* 1997), *LRAT* (Ruiz *et al*., 1999) and *RPE65* (Redmond *et al.,* 1998). Some proteins have a structural role, such as *RP1* (Bowne *et al.,* 1999) and *TULP1* (Banerjee *et al.,* 1998). *NR2E3* (Haider *et al.,* 2000) and *NRL* (Bessant *et al.,* 1999) play a role in transcriptional regulation,

whereas *MERTK* (Gal *et al.,* 2000) is involved in the phagocytosis the photoreceptors by the RPE.

6.1 Traditional and conventional methods for identifying *EYS*

In this project, automated mutation sequencing analysis was performed in an attempt to identify the causative gene for RP25. Up to the point of identification, 60 genes had been collectively screened for mutations. An alternative method, copy number variation (CNV) analysis was then incorporated into the study based on the identification of heterozygous SNPs in a homozygous interval in one arRP affected family. In parallel, both copy number variation (CGH) array analysis and mutation screening identified mutations in novel transcripts which in turn determined *EYS*.

In theory, the traditional method of mutation screening was successful; however it is imperative to highlight that the identification of large deletions (e.g. in the RP73 family) would never have been identified through performing this technique alone. This reveals the importance and possible application of CGH arrays to identify novel disease causing mutations in genes. Another fundamental procedure performed in this project is 5` and 3` RACE. Through performing PCR analysis with bioinformatics predictions of exons alone, the precise composition of *EYS* would not have been wholly accurate. This leads to a very important conclusion that through the combination of traditional molecular techniques and bioinformatics and chip array analysis allow for the most accurate prediction of mutations and novel gene size.

6.2 *EYS* mutations in arRP Spanish families

Prior to the identification of *EYS*, all causative genes for arRP accounted for a minority of cases, it is suggested that *EYS* may be the common causative gene for many arRP cases. We have previously concluded that 10-25% of all Spanish arRP cases are linked to the RP25 interval, the suggestion is that *EYS* is a causative gene for a proportionate number of arRP cases, not only in Spain, but worldwide. This theory is based on the linkage of families from different ethnic backgrounds to the RP25 interval and the identification of further mutations in a second ethnic population (Collin *et al.,* 2008). Unique mutations that span across *EYS* (e.g. exon 17, exon 28 and exon 41) were discovered in five unrelated arRP Spanish families. This leads to the opinion that there is no one individual region for mutations in *EYS*, and that the gene as a whole is crucial for protein function.

6.3 Further ethnic populations linked to *EYS*

Initially, primary focus was on the Spanish families linked to the RP25 interval, for which it was considered that this gene is the principal causative gene for arRP. However, now a second population with novel mutations has been established, leading to the prospect of further populations being linked. To reinforce this, it is important to consider that Pakistani and Chinese families were also originally linked to the RP25 interval (Abd El-Aziz *et al.,* 2007). The affected members of the Chinese families could

still possess mutations in the EYS gene and should be further analysed for mutations, however linkage of the Pakistani family (Figure 1.10) is outside where *EYS* has now been identified. This data, along with the fact that mutations have not yet been identified in several Spanish families may lead to the possibility that there is a second RP25 causative gene in this interval. This occurrence is rare but has previously been reported for other disorders such as Griscelli syndrome, which is a rare autosomal recessive disease characterised by albinism and immunodeficiency (Pastural *et al.,* 1999).

A significant result was that from Collin *et al.,* 2008 who linked two unrelated Dutch families and one Dutch individual to the same RP25 interval. They examined five patients from the two families and discovered the same homozygous nonsense mutation (p.Tyr3156X), whereas in the individual patient, a homozygous frame-shift mutation was identified (p.Pro2238ProfsX16). This, in addition to the information from our studies concludes that there are at least seven novel and independent mutations in the *RP25* gene. The mutations vary from missense mutations, large deletions and null deletions. Moreover, this further reinforces the suggestion that this gene is a major gene for arRP, as a second ethnic population has been linked to the causative gene independently.

Another interesting aspect of this study is the variation in mutations in different ethnic populations, no common mutation has yet been identified across the two disease associated populations. Since mutations in *EYS* are now not only confined to the

Spanish population, it may be possible that different mutations give rise to the disease in different populations? Or possibly the severity and the onset of the disease depend on the type and position of the ensuing mutation in *EYS*.

An intriguing result from this study is the conclusion from data so far is that there is no unique specific hotspot for mutation events in *EYS*, mutations are dispersed across the full length of the gene. This result can be correlated with the major causative gene for dominant Retinitis Pigmentosa, *rhodopsin* (Schuster *et al.,* 2005). In the *rhodopsin* gene, over 200 different mutations across the gene have been identified (Chadderton *et al.,* 2009). It is speculated that *EYS* may be the primary causative gene for arRP, just as *rhodopsin* is for adRP.

6.4 *EYS*-Largest human eye specific gene

We determined that *EYS* is composed of 43 exons and spans 2 MB in size. To date, *dystrophin* is the largest gene in the human genome, covering 2.4 MB and spanning 79 exons (Roberts *et al.,* 1993), located on chromosome Xp21, it is the causative gene for Duchenne Muscular Dystrophy (Hoffman *et al*., 1987). *EYS* is consequently one of the largest genes identified in the human genome. To discover the molecular function of *EYS* could prove to be a difficult and complicated task due to it being such a large gene. The large size of some human disease genes in combination with a large number of distinct disease-causing mutations essentially restricts the identification of the specific molecular alterations underlying disease phenotypes (Van Orsouw *et al.,* 1996).

The full length transcript of *EYS* is retina specific however, smaller transcripts do exist in other tissues, for example a transcript has been identified in human testis tissue between exons 9 and 12 (Collin *et al.,* 2008). An alternatively spliced exon of 63bp was discovered between exon 41 and 42 but does not alter the reading frame of the transcript. Low expressing transcripts were also discovered from exons 41-44 in several tissues, including fetal heart and testis; however these transcripts exist at significantly low levels. Even though these small transcripts exist, it is still plausible to conclude that *EYS* is retinal specific, since the full length gene transcript exists exclusively in the retina.

6.5 *EYS* and Drosophila

The common fruit fly, *Drosophila melanogaster,* has been a central part of genetic research over the past 100 years (Banfi *et al,* 1997). The *Drosophila* genome sequence was published in 2000 and has since been used as a critical resource for research in genetics and medicine. The high degree of conservation of genes from flies through to mammals is vast and in particular the high degree of conservation of genes causing retinal degeneration in flies and mammals is immense when considering the differences in relation to the development and the organization of the eye of flies (Knust 2007) (Table 6.1).

Importantly, there seems to be a strong correlation between the function of several retinal genes in humans and in *Drosophila* for example Rhodopsin, which is involved

in retinal signal transduction of the Drosophila eye (table 6.1). EYS was determined to have an ortholog in *Drosophila*, this being eys/eys shut with the protein name spam/spacemaker. This led to the corresponding human protein name of SPAM. Many other human genes identified as causative retinal degenerative disease genes have orthologs in *Drosophila*, for example mutations in mammalian *crumbs* (*Crb1)* are causative for RP12- and LCA-related blindness in humans (Knust, 2007). In *Drosophila*, crumbs is very important for photoreceptor development (Fan *et al.,* 2003). Another important retinal degeneration gene is *prominin* (PROM), mutations in which cause arRP and macular degeneration in humans (Maw *et al.,* 2000). In *Drosophila*, prominin, like spam localises in the interhabdomeral space. Mutant prominin flies fail to separate their rhabdomeres, leaving them fused together to form a "closed visual system" (Zelhof *et al.,* 2006).

Drosophila spacemaker/spam is an extracellular matrix protein that is located in the interrhabdomeral space (Collin *et al.,* 2008, Zelhof *et al.,* 2008). The generation of the interrhabdomeral space has been a crucial event in the transition of compound eyes from closed to open visual systems. Due to the development of the open visual systems in fly eyes, they have an improved angular sensitivity, which allows the detection of smaller moving objects. Studies have demonstrated that if spam/spacemaker is knocked out of the *Drosophila* eye, the rhabdomeres fuse together leaving in place of the interrhabdomeral space (Zelhof *et al.,* 2006) (figure 6.1). As previously stated, this fusion is also associated with mutant *Drosophila* prominin line. This leads to the conclusion that spam binds to prom in forming the open rhabdomere structure (Zelhof

et al., 2008). Data from *Drosophila* studies reinforces the suggested function of spam as being an extracellular matrix protein essential for rhabdomere separation (figure 6.2). Spacemaker/spam distribution is linked to prominin in *Drosophila*, while in vertebrates PROMININ localises to the outer segment of rod photoreceptor cells as does SPAM as proven through IHC experiments (Cook and Zelhof 2008). Interestingly, *Drosophila* crumbs is expressed at the stalk membrane, this leads to the speculation that there is an interaction between spam, prominin and possibly crumbs which may be correlated with the human gene orthologues.

An intriguing aspect of this study demonstrates that mutated spam does not lead to retinal degeneration in *Drosophila*. Apart from Spam being a necessary component in the formation of the IRS, it also acts in mechanoreceptor neurons (MRNs) in flies, and is an extracellular shield that protects these neurons from environmental stresses (Cook *et al.,* 2008). MRNs are essential for functions such as hearing, flight control and touch sensing, the aberrant function of these neurons leads to uncoordinated movement in flies therefore demonstrating the apparent dual function of Spam in Drosophila.

Table 6.1 Common eye genes of Drosophila and Humans

Drosophila protein	**Human Homolog**	**Retinal disorder**	**Structural characteristics**	**Function in the eye**
Actin	Yes	Unknown	Secreted protein, Egf-like and Lam A	Formation of interrhabdomeral space
Armadilio/β-catenin	Yes	Unknown	B-catenin like repeats	Cell-cell adhesion
Bazooka	Yes	Unknown	Scaffolding protein	Polarity of Photoreceptor cells (PRCs)
Chaoptin	Yes	Unknown	Transmembrane protein	Photoreceptor-specific adhesion molecule
Crumbs	Yes	LCA	Transmembrane protein. EGF-like and Lam A repeats	Elongation of the PRC/rhadomere formation/prevention of light-induced degeneration
DE-Cadherin	Yes	Unknown	Transmembrane protein	Cell-cell adhesion
DPATJ	Yes	Unknown	Scaffolding protein	Elongation of the PRC/rhadomere formation/prevention of light-induced degeneration
Prominin	Yes	arRP	Pentaspan membrane glycoprotein	Formation of the interrhabdomeral space
Rhodopsin	Yes	adRP	G protein-coupled receptor with seven transmembrane domains	Signal transduction; morphogenesis of rhabdomere
Rab11	Yes	Unknown	Monomeric GTPase	Mediate membrane transport to developing rhabdomeres
Spacemaker/spam/Eyes shut	Yes	arRP	Secreted protein, EGF-like and laminin A G-domain-like repeats	Formation of the interrhadomeral space

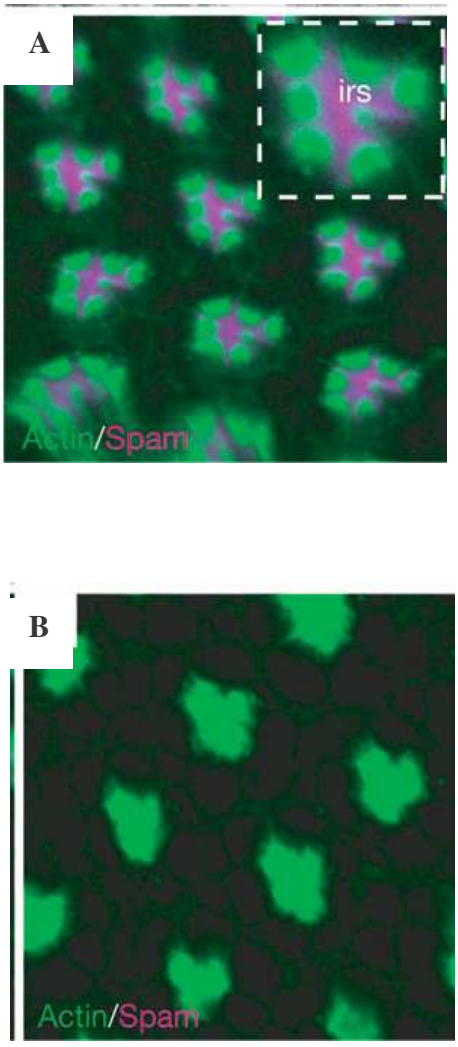


Figure 6.1 Confocal microscopy images of rhabdomeres in a Drosophila eye. **A** Spam wildtype (pink) is evidently located in the IRS (interrhabdomeral space). **B** Spam knockout Drosophila. The absence of Spam is clearly visible through the merging of the rhabdomeres (Images adapted from Zelhof *et al.,* 2006).

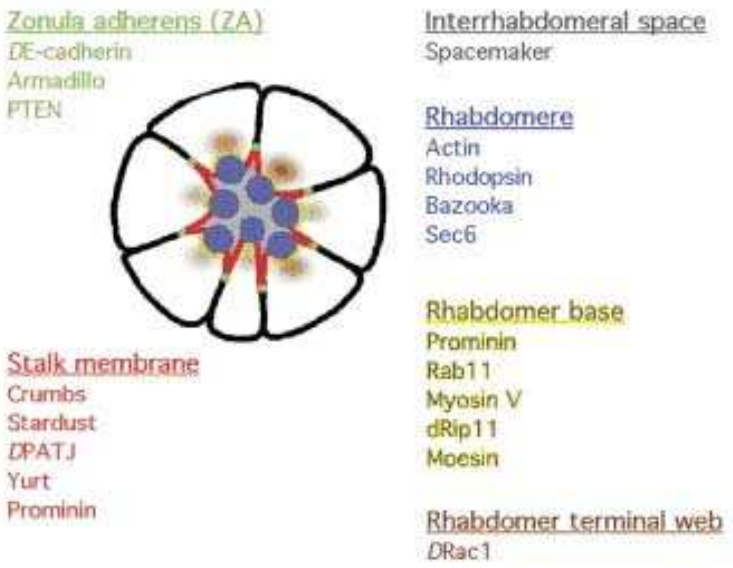


Figure 6.2 Overview of important genes in the structure of the eye of *Drosophila*. Spacemaker/Spam is localised to the interrhabdomeral space. Genes such as prominin and crumbs are located in the neighbouring stalk membrane (Husain *et al*., 2006)

6.6 Orthologues of SPAM

Moving from DNA to protein composition, it is the C-terminus of the SPAM protein that is orthologous to spam in *Drosophila* (figure 4.37). Although the C-terminal amino acids of *Drosophila* spam are not highly conserved in comparison to human SPAM, several residues of this fragment are conserved in vertebrate species, including zebrafish. This indicates that during the evolution of the vertebrate eye, the C-termimus part of SPAM possibly became essential for proper functioning of the entire protein (Collin *et al.,* 2008). However, it is imperative to highlight that the functional antibody for SPAM was raised against a fragment of the N terminus of *EYS,* demonstrating that this region of SPAM is also a fundamental and conserved part of the protein. IHC experiments established this through signal detection in porcine retinal sections using this antibody.

6.7 Absence of SPAM in other species

Through bioinformatics studies it was determined that SPAM is absent from rodent species such as mouse and rat (Section 4.9). Based on the high homology of disease genes between humans and mice, the mouse genome is a commonly used animal model for countless genetic disorders. Many functional and anatomical similarities exist in the eyes of mice and humans, making the mouse an excellent model system for human retinal disorders (Bedell *et al.,* 1997). In this instance the mouse model system would

not be suitable due to the absence of SPAM in the mouse genome. The loss or absence of a gene from humans to mice is a rare occurrence, this is only the fifth human disease gene identified as being absent in the mouse genome. The five genes are as follows and don't seem to follow any particular pattern of expression or function. The genes are *SHOX* (Blaschke *et al.,* 1998), *CETP* (Hogarth *et al.,* 2003), *Fuc-t III* (Huang *et al.,* 2004) and *Kal1* (Soderland *et al.,* 2002).

Using a recently developed phylogenetic approach, Church *et al.,* 2009 identified 15,187 human and mouse genes in 1:1 orthologous relationships, representing 75% of mouse and 80% of human genes. Genes that are associated with Mendelian disease in humans are assumed to be of great importance and function in other species. Some exceptions do occur as in the case of *EYS*, the disruption of *EYS* in rodent lineages was unexpected. One similar example is the human gene *KAL1*, the disruption of this gene in humans leads to an absence or hypoplasia of olfactory bulbs and tracts in patients (Church *et al.,* 2009) however it is absent from the mouse genome. In summary, most of the genes that are essential in humans are also essential in mice, re-enforcing the value of the mouse model for most human genetic diseases. However, a surprisingly high number of genes that are essential in humans are not identified in mice. This suggests that it may not always be easy to draw a parallel between a human disease mutation and a mouse mutant phenotype; this is especially true for *EYS*.

There are several postulations as to why there is no ortholog of *EYS* in the genome of particular species, in the mouse the ortholog is a pseudogene with only very short runs

of nucleotide sequences being identified through BLAST searches. The inability to detect orthologs in mice due to rapid evolutionary pressure has previously been demonstrated with the human gene AU022751 and mouse EG624310 which are orthologous, but only display nonsignificant sequence similarity in database searches (Church *et al.,* 2009). In other species such as cattle and sheep, the absence of *EYS* may be due to incomplete genomic sequencing of the species or the region where the *EYS* ortholog would be mapped to is completely absent.

A recent study by Bekpen *et al.,* 2009 have established that some genes that are essential in humans may have undergone some sort of event to render them as pseudogenes in other species. The *IRGM* gene, which is an immunity-related GTPase, has recently been implicated a as risk factor for Crohn's disease. The study suggests that in the anthropoid ancestors a retrotransposition event occurred disrupting the open reading frame to cause a pseudogene. Millions of years later, in the human-great ape ancestor, the open reading frame became active again. However, most interestingly is that the version of *IRGM* identified in the monkey contains premature stop codons, leading to a pseudogene. Also orangutans, shared a premature stop codon-inducing substitution with monkeys, suggesting ape ancestors once carried a pseudogene similar to that in monkeys. *IRGM* could play an important role in humans; however it is unclear what led to this gene's disuse in the monkey-ape ancestor. It is suggested that another immune mechanism may have evolved around the same time that *IRGM* became a pseudogene, making it redundant and less crucial. This may be correlated to

EYS and that possibly another gene had evolved at the same time that *EYS* became a pseudogene in rodents, making it less crucial and possibly redundant in that lineage.

6.8 Proposed role of SPAM

What is the proposed role for SPAM in human eyes? Based on the function of spam in *Drosophila* and the localisation of SPAM in the outer segments of photoreceptors in porcine retina (Figure 5.8), the initial postulation would be that SPAM provides structure and support to these outer segments. Since Retinitis Pigmentosa is a late onset disorder, that slowly progresses (Tsujikawa *et al.,* 2008), it has been suggested that the degeneration observed in the absence of SPAM could be the end result of structural abnormalities of the outer segments that accumulate over time (Cook and Zelhof, 2008). What about interaction between the human orthologs of prom, eys and crb? *PROM1* is found mainly in rod photoreceptors (Jászai *et al.,* 2007), and from our immunolocalisation studies, it can be assumed that SPAM is also mainly localised to rod outer segments. Therefore it is proposed that there's an interaction between PROM and SPAM in the rod outer segments, leading to disc morphogenesis (Collin *et al.,* 2008). *EYS* may also interact with *CRB1* or *CRB2* and may have a role in Müller cell–photoreceptor cell and photoreceptor cell–photoreceptor cell interactions, as *CRB1* is thought to be required to regulate number and size of Müller glia cell villi (van de Pavert *et al.,* 2007). Another possible function of *EYS* could be that it is restored in the extracellular matrix, also known as the sub retinal space, between the (developing) photoreceptors and RPE. This proposed function is based on the function of spam in

Drosophila eyes, and its importance for the structure of the extracellular matrix in their eyes.

Future Work

The next stage in the investigation of *EYS* will be to perform an intensive mutation screening analysis of *EYS* in arRP panels. This will determine the prevalence of mutations in *EYS* and its expected that novel mutations will be identified in other ethnic arRP populations. Another area of study will be to examine the molecular mechanisms underlying RP25. Through in situ cellular and tissue analysis, the localisation of SPAM has been presented in Y79 cells and porcine retina sections. An important aspect of the future study of SPAM will be to fully characterise the protein which will involve characterising the antibody through western blot analysis and further in-depth IHC and ICC. This will reiterate the data confirmed from IHC performed on porcine retina sections and gives an indication of the expression level of SPAM in human retina and should further confirm the precise localisation of SPAM in human retina. IHC studies will also determine whether SPAM is rod specific in humans, or if there is a level of expression in cone photoreceptors also. Developmental studies of SPAM will also be a critical area of study, through experiments performed on tissue from an early stage developmental source and an adult source of the same species it will be possible to postulate whether SPAM plays a role in controlling or regulating development. Through parallel examination of early stage developmental retinal and mature retina, it will be possible to specify at what stage is SPAM turned on.

As for protein-protein interactions, from the *Drosophila* data it is expected that PROMININ and CRB1 are the protein interactants in humans, however this must be determined using the yeast two hybrid system, through which novel protein interactants may also be identified. The concept of the yeast-two-hybrid technique involves using the GAL4 transcription factor, which has two domains (DNA binding domain and activation domain); the two must come together for transcription to occur. The 'bait' (protein of interest) is expressed as a fusion protein in frame with the DNA binding domain of the GAL4 transcription factor. The 'prey' (potential interactants) is expressed in frame with the activating domain. If the two domains are brought close to one another, by the interaction of the bait and prey, transcription is triggered of a reporter gene, which allows the yeast to grow in selective media lacking a particular nutrient (Coates and Hall, 2003). The function of SPAM will need to be thoroughly investigated and examined.

The secondary and tertiary structure of SPAM will also be required to be studied. Motifs of the secondary structure will reveal substructures such as alpha helices and identify how many motifs are present in a single protein molecule. This will lead onto the identification of the tertiary structure which will reveal how the protein folds and will expose the spatial arrangement of the secondary structures.

Another fascinating aspect of the future of this study will be the evolutionary aspect of *EYS*. Why has *EYS* been disposed of with in particular species and is crucial for

functional eyes in other species? Is there a correlation or connection between these species? Have these genomes been fully sequenced, or has *EYS* become a pseudogene in all non expressing animals? This will require a great amount of effect but should reveal some interesting aspects of eye and retinal gene evolution.

Animal models are a crucial component of biomedical research, both to understand the pathogenesis of human disease and to provide systems for testing new therapies. The mouse has been paramount in modeling human diseases, primarily because of the striking homology between mammalian genomes and the many similarities in aspects spanning from anatomy to cell biology and physiology. The mouse is still a very popular model to use in terms of studying retinal degenerations. Significantly, there are several mouse lines that copy different types of human retinal degeneration (Bovolenta and Cisneros 2009), knockout mice that are commonly used for many forms of retinal degenerations; include the *RDS* and *rpe65* knockout mice. Knock in mouse models are also commonly used for retinal degenerations, the purpose of knock in mice is to determine what the consequences are on the mouse, of having particular mutated genes knocked in. One example of such a model used to study retinal degenerations is the rhodopsin-GFP fusion knock in mouse model. This particular model has been identified to test for arRP and has produced expression patterns that are identical to normal mouse rhodopsin (Chan *et al.,* 2004). One knock-in mouse model has previously been generated in our lab for the *Prpf31* gene, the causative gene for RP11. Due to the large size of *EYS*, it is proposed that a smaller more applicable construct (mini gene) be

prepared for transfection experiments. It will be imperative that all crucial elements of the gene be retained in the miniaturised construct.

Many researchers who are interested in an embryologically and genetically tractable disease model have now turned to zebrafish (*Danio rerio*) as an alternative animal model (Lieschke and Currie 2007*).* In terms of a knock out model for SPAM, it is impossible to utilise the mouse as an animal model, however it will be feasible to construct a knock-in model to determine the effects on the mouse when it possesses the mutated gene. The proposed model for SPAM will be the zebrafish. An ortholog to SPAM has been identified in zebrafish, and the lifespan of the fish allows good scope in terms of retinal degeneration disease models. The symptoms for some arRP cases present mostly in mid to late years and based on the lifespan of the zebrafish, being approximately five years (Gerhard *et al.,* 2002), it is possible to utilise this model over a short period of time. Using zebrafish to model arRP will require a certain level of generalisation; however this system has previously been successfully used as a model organism for several other retinal degenerations. One example of this is choroideremia, which is an X-linked hereditary retinal degeneration caused by mutations in the *REP1* gene. Zebrafish rep1 mutants exhibit degeneration of the RPE and photoreceptors (Krock *et al.,* 2007).

Another method which will be used to investigate *EYS* further is through RNA interference (RNAi), which is a post-transcriptional regulatory system within living cells that helps to control which genes are active and how active they are. Technologies

such as RNAi allow for knock down expression of a gene of interest through using short interfering RNA (siRNA) oligonucleotides complementary to the gene of interest. The siRNA duplexes combine with cellular proteins to form the RNA-induced silencing complex (RISC) which recognises and selectively binds complementary sequences.

The eye is a prime target organ for gene therapy of both inherited and acquired ocular disorders and offers an important model system for gene therapy (Bainbridge *et al.,* 2006). For inherited disorders like RP, several methods can be performed to replace or correct "faulty" genes. In the case of *EYS,* a corrective gene may be introduced into the cell through the use of a recombinant adeno- (adeno-associated) viral vector, in the possibility that the virus would deliver the "normal" gene and replace the "disease" gene. *EYS* could perhaps be a prime target gene for gene therapy due to it being a recessive gene and recessive genes are easier to target because the transfer of a single copy of the gene may be sufficient to correct the loss of function that causes the phenotype. The *RPE65* gene, which is responsible for cases of LCA and arRP, has already been successfully used in gene therapy trials. Trials were first attempted on canine models (Acland *et al.,* 2001) but recently Bainbridge *et al.*, 2008, undertook a groundbreaking study using a rAAV vector containing human *RPE65* human cDNA and the human *RPE65* promoter on three human patients suffering from LCA. Such trials can not be attempted before the function and molecular mechanisms of SPAM are fully understood. The future work for SPAM is immense and very intriguing. To understand the complex molecular mechanisms of the protein may take years. If SPAM

is similar to RHODOPSIN in terms of numbers of mutations, then identifying the diverse roles and functions of each component of the protein will take time and effort. It is likely that we have identified a major gene for recessive RP. This is a novel breakthrough as all previously identified causative genes account for only a minority of all cases of arRP. Mutations in *EYS* have been reported in a number of different ethnic populations, making it a prime candidate for a major arRP gene. Such a case would be immense in terms of patient identification and disease prognosis and prevention. If *EYS* can be incorporated into gene therapy trials the potential for retinal degeneration repair is great. Future research on *EYS* will be an intersting area of study within the retinal research community and in the general wider science world.

REFERENCES

Abdellah Z., Ahmadi A., Aimable M., Ainscough R., Almeida J., (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931-945.

Abd El-Aziz, M. M., Barragan, I., O'Driscoll, C. A., Goodstadt, L., Prigmore, E., Borrego, S., Mena, M., Pieras, J. I., El-Ashry, M. F., Safieh, L. A., *et al.* (2008). EYS, encoding an ortholog of Drosophila spacemaker, is mutated in autosomal recessive retinitis pigmentosa. Nat Genet *40*, 1285-1287.

Abd El-Aziz, M. M., El-Ashry, M. F., Barragan, I., Marcos, I., Borrego, S., Antinolo, G., and Bhattacharya, S. S. (2005). Molecular genetic analysis of two functional candidate genes in the autosomal recessive retinitis pigmentosa, RP25, locus. Curr Eye Res *30*, 1081-1087.

Abd El-Aziz, M. M., El-Ashry, M. F., Chan, W. M., Chong, K. L., Barragan, I., Antinolo, G., Pang, C. P., and Bhattacharya, S. S. (2007). A novel genetic study of Chinese families with autosomal recessive retinitis pigmentosa. Ann Hum Genet *71*, 281-294.

Abd El-Aziz, M. M., Patel, R. J., El-Ashry, M. F., Barragan, I., Marcos, I., Borrego, S., Antinolo, G., and Bhattacharya, S. S. (2006). Exclusion of four candidate genes, KHDRBS2, PTP4A1, KIAA1411 and OGFRL1, as causative of autosomal recessive retinitis pigmentosa. Ophthalmic Res *38*, 19-23.

Acland, G. M., Aguirre, G. D., Ray, J., Zhang, Q., Aleman, T. S., Cideciyan, A. V., Pearce-Kelling, S. E., Anand, V., Zeng, Y., Maguire, A. M., *et al.* (2001). Gene therapy restores vision in a canine model of childhood blindness. Nat Genet *28*, 92-95.

Allikmets, R., Shroyer, N. F., Singh, N., Seddon, J. M., Lewis, R. A., Bernstein, P. S., Peiffer, A., Zabriskie, N. A., Li, Y., Hutchinson, A.*, et al.* (1997). Mutation of the Stargardt disease gene (ABCR) in age-related macular degeneration. Science *277*, 1805-1807.

Atchison, D.A. and Smith, G. (2000). Optics of the human eye. Elsevier Science.

Ayoub, G. S., and Matthews, G. (1992). Substance P modulates calcium current in retinal bipolar neurons. Vis Neurosci *8*, 539-544.

Bader, I., Brandau, O., Achatz, H., Apfelstedt-Sylla, E., Hergersberg, M., Lorenz, B., Wissinger, B., Wittwer, B., Rudolph, G., Meindl, A., and Meitinger, T. (2003). X-linked retinitis pigmentosa: RPGR mutations in most families with definite X linkage and clustering of mutations in a short sequence stretch of exon ORF15. Invest Ophthalmol Vis Sci *44*, 1458-1463.

Bainbridge, J. W., Smith, A. J., Barker, S. S., Robbie, S., Henderson, R., Balaggan, K., Viswanathan, A., Holder, G. E., Stockman, A., Tyler, N.*, et al.* (2008). Effect of gene therapy on visual function in Leber's congenital amaurosis. N Engl J Med *358*, 2231-2239.

Bainbridge, J. W., Tan, M. H., and Ali, R. R. (2006). Gene therapy progress and prospects: the eye. Gene Ther *13*, 1191-1197.

Banerjee, P., Kleyn, P. W., Knowles, J. A., Lewis, C. A., Ross, B. M., Parano, E., Kovats, S. G., Lee, J. J., Penchaszadeh, G. K., Ott, J.*, et al.* (1998). TULP1 mutation in two extended Dominican kindreds with autosomal recessive retinitis pigmentosa. Nat Genet *18*, 177-179.

Banfi, S., Borsani, G., Bulfone, A., and Ballabio, A. (1997). Drosophila-related expressed sequences. Hum Mol Genet *6*, 1745-1753.

Barragan, I., Abd El-Aziz, M. M., Borrego, S., El-Ashry, M. F., O'Driscoll, C., Bhattacharya, S. S., and Antinolo, G. (2008). Linkage validation of RP25 Using the 10K genechip array and further refinement of the locus by new linked families. Ann Hum Genet *72*, 454-462.

Barragan, I., Borrego, S., Abd El-Aziz, M. M., El-Ashry, M. F., Abu-Safieh, L., Bhattacharya, S. S., and Antinolo, G. (2008). Genetic analysis of FAM46A in Spanish families with autosomal recessive retinitis pigmentosa: characterisation of novel VNTRs. Ann Hum Genet *72*, 26-34.

Barragan, I., Marcos, I., Borrego, S., and Antinolo, G. (2005). Molecular analysis of RIM1 in autosomal recessive Retinitis pigmentosa. Ophthalmic Res *37*, 89-93.

Barragan, I., Marcos, I., Borrego, S., and Antinolo, G. (2005). Mutation screening of three candidate genes, ELOVL5, SMAP1 and GLULD1 in autosomal recessive retinitis pigmentosa. Int J Mol Med *16*, 1163-1167.

Bartlett, J., and Stirling, D (2008). A short history of the polymerase chain reaction. Methods in Mol. Biol. *226,* 3-6.

Beckmann, G., Hanke, J., Bork, P., and Reich, J. G. (1998). Merging extracellular domains: fold prediction for laminin G-like and amino-terminal thrombospondin-like modules based on homology to pentraxins. J Mol Biol *275*, 725-730.

Bedell, M. A., Largaespada, D. A., Jenkins, N. A., and Copeland, N. G. (1997). Mouse models of human disease. Part II: recent progress and future directions. Genes Dev *11*, 11-43.

Bekpen, C., Marques-Bonet, T., Alkan, C., Antonacci, F., Leogrande, M. B., Ventura, M., Kidd, J. M., Siswara, P., Howard, J. C., and Eichler, E. E. (2009). Death and resurrection of the human IRGM gene. PLoS Genet *5*, e1000403.

Bessant, D. A., Payne, A. M., Mitton, K. P., Wang, Q. L., Swain, P. K., Plant, C., Bird, A. C., Zack, D. J., Swaroop, A., and Bhattacharya, S. S. (1999). A mutation in NRL is associated with autosomal dominant retinitis pigmentosa. Nat Genet *21*, 355-356.

Biggiogera, M. and Biggiogera, F. F (1989). Ethidium bromide and propidium iodide-PTA staining of nucleic acids at the electron microscopic level. Jour of Histochem and Cytochem. *37,* 1161-1166.

Blaschke, R. J., Monaghan, A. P., Schiller, S., Schechinger, B., Rao, Ercole., Padilla-Nash, H., Reid, T. and Rappold, G. A. (1998) SHOT, a SHOX-related homeobox gene, is implicated in craniofacial, brain, heart, and limb development. PNAS, *95*, 2406-2411.

Botstein, D., White, R.L and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms *32,* 314-331.

Bovolenta, P., and Cisneros, E. (2009). Retinitis pigmentosa: cone photoreceptors starving to death. Nat Neurosci *12*, 5-6.

Bowne, S. J., Daiger, S. P., Hims, M. M., Sohocki, M. M., Malone, K. A., McKie, A. B., Heckenlively, J. R., Birch, D. G., Inglehearn, C. F., Bhattacharya, S. S., *et al.* (1999). Mutations in the RP1 gene causing autosomal dominant retinitis pigmentosa. Hum Mol Genet *8*, 2121-2128.

Bykhovskaya, Y., Mengesha, E., Wang, D., Yang, H., Estivill, X., Shohat, M., and Fischel-Ghodsian, N. (2004). Phenotype of non-syndromic deafness associated with the mitochondrial A1555G mutation is modulated by mitochondrial RNA modifying enzymes MTO1 and GTPBP3. Mol Genet Metab *83*, 199-206.

Carroll, P., Renoncourt, Y., Gayet, O., De Bovis, B., and Alonso, S. (2001). Sorting nexin-14, a gene expressed in motoneurons trapped by an in vitro preselection method. Dev Dyn *221*, 431-442.

Carter, N. P. (2007). Methods and strategies for analyzing copy number variation using DNA microarray. *39*, 16-21.

Chadderton, N., Millington-Ward, S., Palfi, A., O'Reilly, M., Tuohy, G., Humphries, M. M., Li, T., Humphries, P., Kenna, P. F., and Farrar, G. J. (2009). Improved retinal function in a mouse model of dominant retinitis pigmentosa following AAV-delivered gene therapy. Mol Ther *17*, 593-599.

Chakarova, C. F., Papaioannou, M. G., Khanna, H., Lopez, I., Waseem, N., Shah, A., Theis, T., Friedman, J., Maubaret, C., Bujakowska, K.*, et al.* (2007). Mutations in TOPORS cause autosomal dominant retinitis pigmentosa with perivascular retinal pigment epithelium atrophy. Am J Hum Genet *81*, 1098-1103.

Chakravarti, A. (2001). To a future of genetic medicine. Nature *409*, 822-823.

Chan, F., Bradley, A., Wensel, T. G., and Wilson, J. H. (2004). Knock-in human rhodopsin-GFP fusions as mouse models for human disease and targets for gene therapy. Proc Natl Acad Sci U S A *101*, 9109-9114.

Chen, X. N., Korenberg, J. R., Jiang, M., Shen, D., and Fong, H. K. (1996). Localization of the human RGR opsin gene to chromosome 10q23. Hum Genet *97*, 720-722.

Choi, K., Ma, Y., Choi, J-H, and Kim, S. (2005) Platcom: a platform for computational comparative genomics. Bioinformatics.

Church, D. M., Goodstadt, L., Hillier, L. W., Zody, M. C., Goldstein, S., She, X., Bult, C. J., Agarwala, R., Cherry, J. L., DiCuccio, M.*, et al.* (2009). Lineage-specific biology revealed by a finished genome assembly of the mouse. PLoS Biol *7*, e1000112.

Coates, P. J., and Hall, P. A. (2003). The yeast two-hybrid system for identifying protein-protein interactions. J Pathol *199*, 4-7.

Collins, F. S., Green, E. D., Guttmacher, A. E. and Guyer M. S. (2003). A vision for the future of genomics research. Nature. 422.

Collin, R. W., Littink, K. W., Klevering, B. J., van den Born, L. I., Koenekoop, R. K., Zonneveld, M. N., Blokland, E. A., Strom, T. M., Hoyng, C. B., den Hollander, A. I., and Cremers, F. P. (2008). Identification of a 2 Mb human ortholog of Drosophila eyes shut/spacemaker that is mutated in patients with retinitis pigmentosa. Am J Hum Genet *83*, 594-603.

Cook, B., and Zelhof, A. C. (2008). Photoreceptors in evolution and disease. Nat Genet *40*, 1275-1276.

Cremers, F. P., van de Pol, D. J., van Driel, M., den Hollander, A. I., van Haren, F. J., Knoers, N. V., Tijmes, N., Bergen, A. A., Rohrschneider, K., Blankenagel, A.*, et al.* (1998). Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by

splice site mutations in the Stargardt's disease gene ABCR. Hum Mol Genet *7*, 355-362.

Czarny-Ratajczak, M., Lohinival, J., Rogala, P., Kazimierz, K., Perala, M., Cater, L., Spector, L., Kolodzie, L., Seppanen, U., Glazar, R., Krolewski, J., Latos-Bielenska, A. and Ala-Kokko, L. (2001) A Mutation in COL9A1 Causes Multiple Epiphyseal Dysplasia: Further Evidence for Locus Heterogeneity. *69*,969-980.

den Hollander, A. I., Koenekoop, R. K., Mohamed, M. D., Arts, H. H., Boldt, K., Towns, K. V., Sedmak, T., Beer, M., Nagel-Wolfrum, K., McKibbin, M.*, et al.* (2007). Mutations in LCA5, encoding the ciliary protein lebercilin, cause Leber congenital amaurosis. Nat Genet *39*, 889-895.

Dharmaraj, S., Li, Y., Robitaille, J. M., Silva, E., Zhu, D., Mitchell, T. N., Maltby, L. P., Baffoe-Bonnie, A. B., and Maumenee, I. H. (2000). A novel locus for Leber congenital amaurosis maps to chromosome 6q. Am J Hum Genet *66*, 319-326.

Dryja, T. P., Hahn, L. B., Kajiwara, K. (1997). Dominant and digenic mutations in the peripherin/RDS and ROM1 genes in retinitis pigmentosa. Invest Ophthalmol Vis Sci. *38*, 1972-82.

Dryja, T. P., McGee, T. L., Hahn, L. B., Cowley, G. S., Olsson, J. E., Reichel, E., Sandberg, M. A., and Berson, E. L. (1990). Mutations within the rhodopsin gene in patients with autosomal dominant retinitis pigmentosa. N Engl J Med *323*, 1302-1307.

Efron, B., Halloran, E., and Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. Proc Natl Acad Sci U S A *93*, 13429-13434.

Farrar, G. J., Kenna P. F., and Humphries P. (2002). On the genetics of retinitis pigmentosa and on mutation-independent approaches to therapeutic intervention

EMBO J. *21,* 857–864.

Fan, S. S., Chen, M. S., Lin, J. F., Chao, W. T., and Yang, V. C. (2003). Use of gain-of-function study to delineate the roles of crumbs in Drosophila eye development. J Biomed Sci *10*, 766-773.

Fisk HA, Mattison CP, Winey M. (2003) Human Mps1 protein kinase is required for centrosome duplication and normal mitotic progression. Proc Nat Acad Sci. *100.*14875–14880.

Fliegauf, M., Benzing, T. and Omran, H. (2007) Structure and function of the photoreceptor-connecting cilium. Nature Reviews. *8,* 880-893.

Gal, A., Li, Y., Thompson, D. A., Weir, J., Orth, U., Jacobson, S. G., Apfelstedt-Sylla, E., and Vollrath, D. (2000). Mutations in MERTK, the human orthologue of the RCS rat retinal dystrophy gene, cause retinitis pigmentosa. Nat Genet *26*, 270-271.

Gerhard, G. S., Kauffman, E. J., Wang, X., Stewart, R., Moore, J. L., Kasales, C. J., Demidenko, E., and Cheng, K. C. (2002). Life spans and senescent phenotypes in two strains of Zebrafish (Danio rerio). Exp Gerontol *37*, 1055-1068.

Hageman, G. S., Marmor, M. F., Yao, X. Y., and Johnson, L. V. (1995). The interphotoreceptor matrix mediates primate retinal adhesion. Arch Ophthalmol *113*, 655-660.

Haider, N. B., Jacobson, S. G., Cideciyan, A. V., Swiderski, R., Streb, L. M., Searby, C., Beck, G., Hockey, R., Hanna, D. B., Gorman, S.*, et al.* (2000). Mutation of a nuclear receptor gene, NR2E3, causes enhanced S cone syndrome, a disorder of retinal cell fate. Nat Genet *24*, 127-131.

Hamel, C. P. (2007). Cone rod dystrophies. Orphanet J Rare Dis *2*, 7.

Han, W., Yip, S. P., Wang, J., and Yap, M.K.H (2004). Using denaturing HPLC for SNP discovery and genotyping, and extablishing the linkage disequilibrium pattern for the all-trans-retinol dehydrogenase (RDH8) gene. Jour of human genetics *49,* 16-23.

Hilgert, N., Smith, R., Camp, G. (2009). Function and expression pattern of nonsyndromic deafness genes. Current Mol. Med. *9*, 546-564.

Hoffman, E. P., Brown, R. H., Jr., and Kunkel, L. M. (1987). Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell *51*, 919-928.

Hogarth, C. A., Roy, A. and Ebert, D. L. (2003) Genomic evidence for the absence of a functional cholesteryl ester transfer protein gene in mice and rats. Comp. Biochem. *135,* 219-229.

Huang, H., Winter, E., Huajun, W., Weinstock, K., Heming, X., Goodstadt, L., Stenson, P., Cooper, D. N., Smith, D., Mar, M., Ponting, C. and Fechtel, K (2004) Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. Genomes Biol *5* R47.

Huang, S. H., Pittler, S. J., Huang, X., Oliveira, L., Berson, E. L., and Dryja, T. P. (1995). Autosomal recessive retinitis pigmentosa caused by mutations in the alpha subunit of rod cGMP phosphodiesterase. Nat Genet *11*, 468-471.

Husain, N., Pellikka, M., Hong, H., Klimentova, T., Choe, K. M., Clandinin, T. R., and Tepass, U. (2006). The agrin/perlecan-related protein eyes shut is essential for epithelial lumen formation in the Drosophila retina. Dev Cell *11*, 483-493.

Ionita-Laza, I., Perry, G. H., Raby, B. A., Klanderman, B., Lee, C., Laird, N. M., Weiss, S. T., and Lange, C. (2008). On the analysis of copy-number variations in genome-wide association studies: a translation of the family-based association test. Genet Epidemiol *32*, 273-284.

Inglehearn, C.F. (1998). Molecular genetics of human retinal dystrophies. Eye. *12,* 571-579.

Jain, P., Garibaldi, J. M., and Hirst, J. D. (2009). Supervised machine learning algorithms for protein structure classification. Comput Biol Chem *33*, 216-223.

Jaszai, J., Janich, P., Farkas, L., Fargeas, C., Huttner, W.B. and Corbeil, D (2007). Differential expression of prominin-1 and prominin-2 in major cephalic exoncrine glands of adult mice. Histochem and cell biol *128,* 409-419.

Jin, Z. B., Liu, X. Q., Hayakawa, M., Murakami, A., and Nao-i, N. (2006). Mutational analysis of RPGR and RP2 genes in Japanese patients with retinitis pigmentosa: identification of four mutations. Mol Vis *12*, 1167-1174.

Johnson, S., Halford, S., Morris, A. G., Patel, R. J., Wilkie, S. E., Hardcastle, A. J., Moore, A. T., Zhang, K., and Hunt, D. M. (2003). Genomic organisation and alternative splicing of human RIM1, a gene implicated in autosomal dominant cone-rod dystrophy (CORD7). Genomics *81*, 304-314.

Kaplan, J., Gerber, S., Larget-Piet, D., Rozet, J. M., Dollfus, H., Dufier, J. L., Odent, S., Postel-Vinay, A., Janin, N., Briard, M. L., and et al. (1993). A gene for Stargardt's disease (fundus flavimaculatus) maps to the short arm of chromosome 1. Nat Genet *5*, 308-311.

Katsanis, N., Lupski, J., and Beales, P. (2001) Exploring the molecular basis of Bardet-Biedl syndrome. Human Mol Gen *10,* 2293-2299.

Katz, M.L. and Gao, C.L (1995) Vitamin A incorporation into lipofuscin-like inclusions in retinal pigment epithelium. Mech Ageing Dev. *84,* 29-38.

Khaleduzzaman, M., Sumiyoshi, H., Yasuyoshi, U., Inoguchi, K., Ninomiya, Y and Yoshioka, H.(1997). Structure of the Human Type XIX Collagen (COL19A1) Gene, Which Suggests It Has Arisen from an Ancestor Gene of the FACIT Family. Genomics *45*, 304-312.

Kee, H. J., Koh, J. T., Kim, M. Y., Ahn, K. Y., Kim, J. K., Bae, C. S., Park, S. S., and Kim, K. K. (2002). Expression of brain-specific angiogenesis inhibitor 2 (BAI2) in normal and ischemic brain: involvement of BAI2 in the ischemia-induced brain angiogenesis. J Cereb Blood Flow Metab *22*, 1054-1067.

Khaliq, S., Hameed, A., Ismail, M., Mehdi, S. Q., Bessant, D. A., Payne, A. M., and Bhattacharya, S. S. (1999). Refinement of the locus for autosomal recessive Retinitis pigmentosa (RP25) linked to chromosome 6q in a family of Pakistani origin. Am J Hum Genet *65*, 571-574.

Kim, K. I., Baek, S. H., Jeon, Y. J., Nishimori, S., Suzuki, T., Uchida, S., Shimbara, N., Saitoh, H., Tanaka, K., and Chung, C. H. (2000). A new SUMO-1-specific protease, SUSP1, that is highly expressed in reproductive organs. J Biol Chem *275*, 14102-14106.

Kim, S. K., Suh, M. R., Yoon, H. S., Lee, J. B., Oh, S. K., Moon, S. Y., Moon, S. H., Lee, J. Y., Hwang, J. H., Cho, W. J., and Kim, K. S. (2005). Identification of developmental pluripotency associated 5 expression in human pluripotent stem cells. Stem Cells *23*, 458-462.

Kisselev, O. G. (2005). Focus on molecules: rhodopsin. Exp Eye Res *81*, 366-367.

Koenekoop, R.K. (2009). Why do cone photoreceptors die in rod-specific forms of retinal degenerations? Ophthal Gen. *30,* 152-154.

Knust, E. (2007). Photoreceptor morphogenesis and retinal degeneration: lessons from Drosophila. Curr Opin Neurobiol *17*, 541-547.

Kolb, H., Linberg, K. A., and Fisher, S. K. (1992). Neurons of the human retina: a Golgi study. J Comp Neurol *318*, 147-187.

Krock, B. L., Bilotta, J., and Perkins, B. D. (2007). Noncell-autonomous photoreceptor degeneration in a zebrafish model of choroideremia. Proc Natl Acad Sci U S A *104*, 4600-4605.

Krull. (2001). Segmental organization of neural crest migration. Mech of Devel. *105*. 37-45.

LaFramboise, T., Winckler, W. and Thomas R. K. (2009). A flexible rank-based framework for detecting copy number abberations from array data. Bioinformatics. *25*, 722-728.

Lamb, T. D., and Pugh, E. N., Jr. (2006). Phototransduction, dark adaptation, and rhodopsin regeneration the proctor lecture. Invest Ophthalmol Vis Sci *47*, 5137-5152.

Larade, K., Jiang, Z., Zhang, Y., Wang, W., Bonner-Weir, S., Zhu, H., and Bunn, H. F. (2008). Loss of Ncb5or results in impaired fatty acid desaturation, lipoatrophy, and diabetes. J Biol Chem *283*, 29285-29291.

Leonard, A. E., Bobik, E. G., Dorado, J., Kroeger, P. E., Chuang, L. T., Thurmond, J. M., Parker-Barnes, J. M., Das, T., Huang, Y. S., and Mukerji, P. (2000). Cloning of a human cDNA encoding a novel enzyme involved in the elongation of long-chain polyunsaturated fatty acids. Biochem J *350 Pt 3*, 765-770.

Liang, Y., Fotiadis, D., Filipek, S., Saperstein, D. A., Palczewski, K., and Engel, A. (2003). Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. J Biol Chem *278*, 21655-21662.

Li, Y., Marcos, I., Borrego, S., Yu, Z., Zhang, K., and Antinolo, G. (2001). Evaluation of the ELOVL4 gene in families with retinitis pigmentosa linked to the RP25 locus. J Med Genet *38*, 478-480.

Lieschke, G. J., and Currie, P. D. (2007). Animal models of human disease: zebrafish swim into view. Nat Rev Genet *8*, 353-367.

Liu, Q., Zhou, J., Daiger, S. P., Farber, D. B., Heckenlively, J. R., Smith, J. E., Sullivan, L. S., Zuo, J., Milam, A. H., and Pierce, E. A. (2002). Identification and subcellular localization of the RP1 protein in human and mouse photoreceptors. Invest Ophthalmol Vis Sci *43*, 22-32.

Luo, D-G., Xue, T., King-Wai Y. (2008). How vision begins: An odyssey
PNAS *105*.9855-9862.

Marcos, I., Borrego, S., and Antinolo, G. (2003). Molecular cloning and characterization of human RAB23, a member of the group of Rab GTPases. Int J Mol Med *12*, 983-987.

Marcos, I., Galan, J. J., Borrego, S., and Antinolo, G. (2002). Cloning, characterization, and chromosome mapping of the human GlcAT-S gene. J Hum Genet *47*, 677-680.

Marcos, I., Ruiz, A., Blaschak, C. J., Borrego, S., Cutting, G. R., and Antinolo, G. (2000). Mutation analysis of GABRR1 and GABRR2 in autosomal recessive retinitis pigmentosa. J Med Genet *37*, E5.

Martinez-Mir, A., Paloma, E., Allikmets, R., Ayuso, C., del Rio, T., Dean, M., Vilageliu, L., Gonzalez-Duarte, R., and Balcells, S. (1998). Retinitis pigmentosa caused by a homozygous mutation in the Stargardt disease gene ABCR. Nat Genet *18*, 11-12.

Matic, I., van Hagen, M., Schimmel, J., Macek, B., Ogg, S. C., Tatham, M. H., Hay, R. T., Lamond, A. I., Mann, M., and Vertegaal, A. C. (2008). In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy. Mol Cell Proteomics *7*, 132-144.

Maw, M. A., Corbeil, D., Koch, J., Hellwig, A., Wilson-Wheeler, J. C., Bridges, R. J., Kumaramanickavel, G., John, S., Nancarrow, D., Roper, K.*, et al.* (2000). A frameshift mutation in prominin (mouse)-like 1 causes human retinal degeneration. Hum Mol Genet *9*, 27-34.

Mazzoccoa, M., Maffeia, M., Egeoa, A., Verganoa, A., Arrigob, P., Di Lisic, R., Ghiottod, F. and Scartezzini, P. (2002). The identification of a novel human homologue of the SH3 binding glutamic acid-rich (SH3BGR) gene establishes a new family of highly conserved small proteins related to Thioredoxin Superfamily. Gene *291*, 233-239.

McCarthy, J. J., and Hilfiker, R. (2000). The use of single-nucleotide polymorphism maps in pharmacogenomics. Nat. Biotechnol *18,* 506-510.

Lindberg RA, Fischer WH, Hunter T. Characterization of a human protein threonine kinase isolated by screening an expression library with antibodies to phosphotyrosine. Oncogene. 1993;8:351–359.

Molday, R. S., and MacKenzie, D. (1983). Monoclonal antibodies to rhodopsin: characterization, cross-reactivity, and application as structural probes. Biochemistry *22*, 653-660.

Naeem, A, Saleemuddin, M., Khan, Hasan. (2007). Glycoprotein Targeting and Other Applications of Lectins in Biotechnology. Current Protein and Peptide Science, *8*, 261-271(11).

Nakayama, M., Nakajima, D., Nagase, T., Nomura, N., Seki, N., and Ohara, O. (1998). Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening. Genomics *51*, 27-34.

Osorio, D. (2007). Spam and the evolution of the fly's eye. Bioessays *29*, 111-115.

Pastural, E., Ersoy, F., Yalma, N., Wulffraat, N., Grillo, E., Ozkinay, F., Tezcan, I., Gedikogula, G., Philippe, n., Fischer, A. and Sainte Basille G (1999). Two genes are responsible for Griscelli syndrome at the same 15q21 locus. Genomics *63,* 299-306.

Peake, I. (1989). The Polymerase chain reaction. J Clin Pathol. *42*, 676-6.

Pow, D. V., Sullivan, R., and Scott H (2003). Antibody production and immunocytochemical localization of amino acid transporters. Methods Mol Biol. *227,* 213-44.

Ramos-Vara, J.A. (2005). Technical aspects of immunohistochemistry. Vet Pathol. *42,* 405-406.

Redmond, T. M., Yu, S., Lee, E., Bok, D., Hamasaki, D., Chen, N., Goletz, P., Ma, J. X., Crouch, R. K., and Pfeifer, K. (1998). Rpe65 is necessary for production of 11-cis-vitamin A in the retinal visual cycle. Nat Genet *20*, 344-351.

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H., Shapero, M. H., Carson, A. R., Chen, W.*, et al.* (2006). Global variation in copy number in the human genome. Nature *444*, 444-454.

Reinhard, J., Messias, A., Dietz, K., Mackeben, M., Lakmann, R., Scholl, H. P., Apfelstedt-Sylla, E., Weber, B. H., Seeliger, M. W., Zrenner, E., and Trauzettel-Klosinski, S. (2007). Quantifying fixation in patients with Stargardt disease. Vision Res *47*, 2076-2085.

Reme, C.E. and Wenzel, A. (2003). The dangers of seeing in the dark. Nat Gen. *35,* 115-116.

Richard, M., Roepman, R., Aartsen, W. M., van Rossum, A. G., den Hollander, A. I., Knust, E., Wijnholds, J., and Cremers, F. P. (2006). Towards understanding CRUMBS function in retinal dystrophies. Hum Mol Genet *15 Spec No 2*, R235-243.

Rivolta, C., Sharon, D., DeAngelis, M. M., and Dryja, T. P. (2002). Retinitis pigmentosa and allied diseases: numerous diseases, genes, and inheritance patterns. Hum Mol Genet *11*, 1219-1227.

Roberts, R. G., Coffey, A. J., Bobrow, M., and Bentley, D. R. (1993). Exon structure of the human dystrophin gene. Genomics *16*, 536-538.

Robinson, J.M. (2001). Antigen retrieval in cells and tissues: enhancement with sodium dodecyl sulfate. Histochem and cell biol. *116,* 119-130.

Rosenfeld, P. J., Cowley, G. S., McGee, T. L., Sandberg, M. A., Berson, E. L., and Dryja, T. P. (1992). A null mutation in the rhodopsin gene causes rod photoreceptor dysfunction and autosomal recessive retinitis pigmentosa. Nat Genet *1*, 209-213.

Ruiz, A., Borrego, S., Marcos, I., and Antinolo, G. (1998). A major locus for autosomal recessive retinitis pigmentosa on 6q, determined by homozygosity mapping of chromosomal regions that contain gamma-aminobutyric acid-receptor clusters. Am J Hum Genet *62*, 1452-1459.

Ruiz, A., Winston, A., Lim, Y. H., Gilbert, B. A., Rando, R. R., and Bok, D. (1999). Molecular and biochemical characterization of lecithin retinol acyltransferase. J Biol Chem *274*, 3834-3841.

Sasaki, T., Fassler, R., and Hohenester, E. (2004). Laminin: the crux of basement membrane assembly. J Cell Biol *164*, 959-963.

Schraermeyer, U., and Heimann, K. (1999). Current understanding on the role of retinal pigment epithelium and its pigmentation. Pigment Cell Res *12*, 219-236.

Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., and Bork, P. (2000). SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res *28*, 231-234.

Schuster, A., Weisschuh, N., Jagle, H., Besch, D., Janecke, A. R., Zierler, H., Tippmann, S., Zrenner, E., and Wissinger, B. (2005). Novel rhodopsin mutations and genotype-phenotype correlation in patients with autosomal dominant retinitis pigmentosa. Br J Ophthalmol *89*, 1258-1264.

Smith, A. J., Bainbridge J. W., Ali R.R. (2009). Prospects for retinal gene replacement therapy. Trends in Genetics *25*, 156-165.

Smith, G. D. and Ebrahim, S. (2004). Mendelian randomization: prospects, potentials, and limitations. Int. J. Epidemiol. *33*. 30-42.

Söderlund, D., Canto, P. and Méndez, J. P. (2002). Identification of Three Novel Mutations in the KAL1 Gene in Patients with Kallmann Syndrome. The Journal of Clin. Endocrinology & Metab. *87,* 2589-2592.

Sumiyoshi, H., Laub, F., Yoshioka, H., and Ramirez, F. (2001). Embryonic expression of type XIX collagen is transient and confined to muscle cells. Dev Dyn *220*, 155-162.

Tien, L., Rayborn, M. E., and Hollyfield, J. G. (1992). Characterization of the interphotoreceptor matrix surrounding rod photoreceptors in the human retina. Exp Eye Res *55*, 297-306.

Travis, G. H., Golczak, M., Moise, A. R., and Palczewski, K. (2007). Diseases caused by defects in the visual cycle: retinoids as potential therapeutic agents. Annu Rev Pharmacol Toxicol *47*, 469-512.

Tsokos, M., Kyritsis, A. P., Chader, G. J., and Triche, T. J. (1986). Differentiation of human retinoblastoma in vitro into cell types with characteristics observed in embryonal or mature retina. Am J Pathol *123*, 542-552.

Tsujikawa, M., Wada, Y., Sukegawa, M., Sawa, M., Gomi, F., Nishida, K., and Tano, Y. (2008). Age at onset curves of retinitis pigmentosa. Arch Ophthalmol *126*, 337-340.

Van Campa, G., Snoeckx, R., L, Hilgerta, N., van den Endea, J., Fukuokae, H., Wagatsumae, M., Suzukiw, H., Smetsb, E., Vanhoenackerc, F., Declaud, F., Van De

Heyningd, P. and Usamie, S. (2006). A New Autosomal Recessive Form of Stickler Syndrome Is Caused by a Mutation in the COL9A1 Gene. Volume *79,* 449-457.

van de Pavert, S. A., Sanz, A. S., Aartsen, W. M., Vos, R. M., Versteeg, I., Beck, S. C., Klooster, J., Seeliger, M. W., and Wijnholds, J. (2007). Crb1 is a determinant of retinal apical Muller glia cell features. Glia *55*, 1486-1497.

Venter, C. V., et al (2001). The Sequence of the Human Genome. Science 291, 1304-1351.

Wassle, H., and Boycott, B. B. (1991). Functional architecture of the mammalian retina. Physiol Rev *71*, 447-480.

Weinmaster, G., Roberts, V. J., and Lemke, G. (1991). A homolog of Drosophila Notch expressed during mammalian development. Development *113*, 199-205.

Whitehouse, D. B., Putt, W., Lovegrove, J. U., Morrison, K., Hollyoake, M., Fox, M. F., Hopkinson, D. A., and Edwards, Y. H. (1992). Phosphoglucomutase 1: complete human and rabbit mRNA sequences and direct mapping of this highly polymorphic marker on human chromosome 1. Proc Natl Acad Sci U S A *89*, 411-415.

Williams, D. S. (2007). Usher syndrome: animal models, retinal function of Usher proteins, and prospects for gene therapy. Vision Research. *48*, 433-441.

Yang, X., Deignan, J. L., Qi, H., Zhu, J., Qian, S., Zhong, J., Torosyan, G., Majid, S., Falkard, B., Kleinhanz, R. R.*, et al.* (2009). Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. Nat Genet *41*, 415-423.

Yaphe, W. (1984). Properties of gracilaria agars. Hydrobiologia *116*, 171-174.

Zelhof, A. C., Hardy, R. W., Becker, A., and Zuker, C. S. (2006). Transforming the architecture of compound eyes. Nature *443*, 696-699.

Zhang, K., Kniazeva, M., Han, M., Li, W., Yu, Z., Yang, Z., Li, Y., Metzker, M. L., Allikmets, R., Zack, D. J., *et al.* (2001). A 5-bp deletion in ELOVL4 is associated with two related forms of autosomal dominant macular dystrophy. Nat Genet *27*, 89-93.

Zhu, H., and Zhou, J. N. (2008). SUMO1 enhances 17-beta estradiol's effect on CRH promoter activation through estrogen receptors. Neuro Endocrinol Lett *29*, 230-234.

Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R., and Siebert, P. D. (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques *30*, 892-897.

Worldwide web resources

http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi

http://searchlauncher.bcm.tmc.edu/

http://www.sigmaaldrich.com/united-kingdom.html

http://commons.wikimedia.org/wiki/File:Immunohistochemistry.jpg

http://blast.ncbi.nlm.nih.gov/Blast.cgi

http://www.expasy.ch/tools/

http://genome.ucsc.edu.

http://smart.embl-heidelberg.de/

http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?=Proteins&PROGRAM=blastp&BLAST_PROGRAMS=blastp&_TYPE=BlastSearch&SHOW_DEFAULTS=on

http://www.ensembl.org/index.html

http://www.nimr.mrc.ac.uk/molneurobiol/salecker/eyesection/

Supplementary Data

EYS cDNA and UTR (OMIM 612424)

(see also Table 4.1)

agaaacaaagtgatacacgctaggcatgtccagacattaaccgtaatcagaagcctcctagggttgcaagcctccttggct
ggttgggctttgtgcctcactcttctcttacacctggggagtcaagtctggatcctctgctccatatttatttagcaaataaaaca
ctgatctcaaacagcaacggaacctgtgtacttgagttttaaaggtgttttttgaaaactaatcagaagaggaatgcttatctcc
caagaaaattgaaaccatctgtaaaagaatcttggctgtgctgtgttttgattagattgaagcttggtgtaaagagaaaacaca
aaagaccgcaagaaacaaagtgatacacgctaggcatgtccagacattaaccgttaaccgtaatcagaagcctcctaggg
ttgcaagcctccttggctaatggttgggctttgtgcctcactcttctcttacacctggggagtcaagtctggatcctctgctcca
tatttatttagcaaataaaacactgatctcaaacagcaacggaacctgtgtacttgagttttaaaggtgttttttgaaaactaatc
agaagaggaatgcttatctcccaagaaaattgaaaccatctgtaaaagaatcttggctgtgctgtgttttgattagattgaagc
ttggtgtaaagagaaaacacaaaagaccgcagaatatttacaacatttgaattcctgtttctccatcaaagcacattgtcatca
aagctactttgggtatagaaaaaaggtcaattccaattcccaggaatccttaaccacaactttacttggaaatttccggtaata
cttaaaacaccattttgcagcaattctttatgataaactgtaaaatacagctacccgaaaATGACTGACAAATC
AATCGTCATTCTGAGCCTGATGGTTTTTCACAGCTCTTTCATAAATGGAAA
AACATGTAGACGGCAATTGGTGGAAGAATGGCATCCACAACCCTCATCAT
ATGTGGTAAATTGGACACTAACAGAAAACATCTGCTTGGACTTCTACAGA
GATTGCTGGTTTTTGGGTGTAAACACTAAAATAGATACTTCAGGCAATCA
AGCTGTTCCCCAGATTTGCCCTTTGCAAATTCAATTAGGAGATATCCTTGT
AATTTCTTCTGAACCATCTCTTCAATTTCCAGAAATAAATTTGATGAATGT
TTCTGAAACATCTTTCGTTGGCTGTGTGCAAAATACCACAACGGAAGATC
AGTTACTTTTTGGCTGCAGACTAAAAGGAATGCACACTGTTAATTCTAAG
TGGCTGAGTGTTGGGACACATTATTTTATCACAGTTATGGCAAGTGGTCC
ATCACCTTGTCCACTGGGACTTCGACTAAATGTGACAGTGAAACAGCAGT
TCTGCCAGGAATCTCTGAGTTCAGAATTTTGCTCTGGTCATGGTAAATGTC
TTAGTGAAGCTTGGAGCAAGACATATAGCTGCCATTGCCAGCCTCCATTT
TCTGGAAAATACTGCCAGGAACTTGATGCATGTTCTTTTAAACCATGTAA
AAATAATGGCAGTTGCATTAATAAAAGAGAAAATTGGGATGAGCAAGCA
TATGAATGTGTCTGTCACCCACCATTTACAGGAAAGAATTGCTCAGAAAT
AATTGGCCAGTGTCAACCACATGTCTGTTTCCATGGAAACTGCAGCAATA
TTACTTCAAATAGTTTCATTTGTGAATGTGATGAGCAATTTTCAGGTCCAT
TCTGTGAGGTGTCAGCAAAACCTTGTGTTTCTCTGCTTTTTTGGAAAAGAG
GAATTTGCCCAAATAGCAGTTCTGCTTATACTTATGAATGCCCAAAAGGA
TCTTCCAGCCAAAATGGTGAAACTGATGTCAGTGAGTTTTCATTAGTACC
ATGTCAGAATGGTACTGACTGCATCAAAATATCAAATGATGTTATGTGCA
TCTGTTCACCAATATTTACAGATTTGCTTTGTAAGAGCATTCAAACATCAT
GTGAGTCATTTCCTTTGAGGAATAATGCTACATGTAAGAAATGTGAGAAA
GATTATCCTTGCAGCTGTATATCAGGATTTACTGAAAAAAACTGTGAGAA

AGCAATTGACCACTGTAAACTGCTCAGCATCAACTGTCTGAATGAAGAAT
GGTGTTTCAATATAATTGGAAGATTCAAATATGTATGCATTCCAGGGTGC
ACAAAAAATCCATGTTGGTTTTTGAAGAATGTTTACCTAATTCATCAACAC
CTCTGCTACTGTGGAGTCACCTTCCATGGTATTTGCCAAGATAAAGGTCCT
GCTCAATTTGAATATGTGTGGCAATTGGGATTTGCAGGATCTGAAGGCGA
AAAGTGCCAAGGGGTTATTGATGCCTATTTCTTTCTGGCTGCAAACTGCAC
TGAAGATGCAACCTATGTGAACGATCCTGAAGATAATAATTCTTCATGTT
GGTTCCCACATGAAGGCACAAAAGAGATTTGTGCAAATGGATGCAGTTGT
TTGAGTGAAGAAGACAGTCAGGAATATCGGTATCTATGTTTTCTCAGATG
GGCTGGCAACATGTATCTGGAAAATACAACTGATGATCAAGAAAATGAG
TGTCAACATGAAGCTGTTTGTAAAGATGAAATTAATAGACCCAGATGCAG
CTGTTCTCTTAGTTACATTGGCAGATTGTGTGTTGTCAATGTTGACTATTG
CTTAGGGAACCACAGTATATCAGTGCATGGCCTCTGCCTGGCCCTTTCGC
ACAATTGTAACTGTAGCGGTCTGCAAAGATATGAAAGGAACATCTGTGAG
ATAGATACTGAAGACTGCAAATCTGCGTCCTGCAAAAATGGAACAACTAG
TACACATTTAAGGGGATATTTCTTCCGCAAGTGTGTCCCAGGATTTAAAG
GTACGCAATGTGAAATTGATATAGATGAGTGTGCTTCACATCCCTGCAAA
AATGGAGCCACCTGCATTGACCAACCTGGTAATTACTTCTGCCAGTGTGT
GCCTCCATTTAAAGTGGTTGATGGCTTTTCCTGCTTATGCAATCCAGGCTA
TGTTGGGATAAGATGTGAACAGGACATTGATGACTGCATCCTGAATGCCT
GTGAGCACAATTCTACCTGCAAAGACCTGCATCTCAGCTACCAGTGTGTG
TGCCTATCTGATTGGGAAGGAAATTTTTGTGAACAAGAATCCAATGAGTG
TAAAATGAATCCTTGCAAGAACAATTCCACCTGTACTGACCTTTACAAAA
GCTATCGATGTGAGTGTACATCTGGATGGACTGGACAGAACTGTAGTGAA
GAAATAAATGAATGCGACTCTGATCCATGCATGAATGGAGGTCTTTGTCA
TGAATCTACCATCCCTGGACAATTTGTATGTCTGTGCCCACCCCTTTATAC
TGGACAATTTTGCCACCAACGCTATAACCTTTGTGACCTACTTCATAACCC
TTGCAGAAACAACTCAACATGCTTAGCTTTGGTAGACGCAAATCAGCATT
GTATTTGTAGAGAAGAATTTGAAGGTAAAAACTGTGAAATTGATGTGAAA
GACTGCCTCTTCCTTTCCTGCCAGGATTATGGTGACTGTGAAGATATGGTC
AACAATTTCAGGTGTATTTGCAGACCTGGGTTTTCTGGATCTCTGTGTGAA
ATTGAAATTAATGAATGTTCCTCTGAACCTTGCAAAAATAATGGAACATG
TGTGGATCTGACAAACAGATTTTTTTGTAATTGTGAACCTGAGTACCATGG
GCCCTTCTGTGAACTTGATGTAAATAAATGTAAAATCTCACCTTGTCTAGA
TGAAGAAAATTGTGTCTACAGGACTGATGGATACAACTGCCTCTGTGCCC
CTGGTTATACAGGCATCAACTGTGAAATAAATCTAGATGAATGCCTATCA
GAGCCCTGTCTCCATGATGGAGTTTGTATCGATGGCATCAATCATTATACC
TGTGACTGCAAGAGTGGGTTTTTTGGAACACACTGTGAAACAAACGCCAA
TGATTGCCTTTCAAATCCTTGTCTACATGGAAGGTGCACAGAACTTATTAA
TGAATATCCATGTTCATGTGATGCAGATGGGACTAGCACACAATGTAAGA
TCAAAATTAATGACTGCACATCAATCCCTTGTATGAATGAAGGCTTCTGTC
AGAAGTCAGCACATGGATTTACTTGCATTTGCCCACGTGGATACACTGGT
GCATACTGTGAAAAAAGCATTGATAATTGTGCTGAGCCTGAACTTAATTC
AGTCATCTGTCTTAATGGAGGGATCTGTGTTGATGGGCCTGGACATACTTT
TGACTGCAGATGTCTTCCTGGATTTTCTGGTCAATTTTGTGAAATTAATAT

AAATGAATGCTCTTCATCACCATGTCTACATGGTGCAGACTGTGAAGATC
ACATCAATGGATATGTTTGCAAATGCCAACCAGGATGGTCTGGACACCAC
TGTGAGAATGAGCTTGAGTGCATTCCCAACTCATGTGTTCATGAACTCTGC
ATGGAGAATGAACCTGGCTCGACATGTTTATGCACACCTGGATTTATGAC
CTGCTCCATTGGGCTTCTTTGTGGTGATGAAATAAGGAGAATTACCTGTTT
AACTCCCATCTTTCAAAGAACAGATCCCATTTCCACACAGACATATACAA
TTCCCCCTTCTGAGACTTTGGTCAGCAGCTTTCCATCTATAAAGGCTACTA
GAATACCAGCCATAATGGACACTTACCCAGTTGATCAAGGTCCAAAACAG
ACAGGCATTGTCAAGCACGACATTCTCCCAACCACTGGTTTGGCAACATT
AAGAATTAGCACACCCTTGGAAAGCTACTTACTCCAAGAACTGATTGTCA
CTAGAGAGCTTTCAGCAAAACACAGTCTTCTTTCTTCCGCAGATGTTTCCT
CTTCTCGATTCCTGAATTTTGGTATTCGTGACCCAGCACAAATTGTCCAGG
ACAAAACATCGGTATCACATATGCCTATTCGAACTTCAGCAGCCACACTA
GGTTTCTTTTTTCCTGATAGGAGAGCAAGGACCCCATTTATCATGTCTTCC
TTAATGTCAGATTTTATTTTTCCTACACAATCTTTATTATTTGAGAACTGTC
AGACTGTTGCTTTATCTGCTACCCCAACGACTTCAGTAATTAGAAGCATTC
CAGGGGCTGATATTGAGCTAAACAGGCAGTCATTACTCTCCCGTGGATTC
CTGCTTATAGCTGCCTCCATAAGTGCAACTCCAGTTGTCTCTAGGGGGGCT
CAAGAGGATATTGAAGAATATTCAGCTGATTCTTTAATTTCAAGAAGAGA
GCACTGGAGATTGCTCAGCCCCTCGATGTCTCCCATTTTTCCTGCTAAGGT
AATTATATCTAAACAGGTAACCATCTTAAACTCATCAGCTCTGCACCGGTT
CAGTACAAAAGCCTTCAATCCCAGTGAATATCAGGCTATTACTGAGGCTT
CAAGCAACCAGAGACTCACAAACATCAAATCACAGGCTGCTGATTCTTTA
AGAGAATTAAGCCAAACATGTGCAACATGTTCTATGACTGAAATAAAATC
CTCTCGTGAATTCTCAGATCAAGTTTTGCATAGCAAACAGTCCCACTTTTA
TGAGACATTCTGGATGAACTCAGCGATATTAGCCAGCTGGTATGCACTAA
TGGGAGCTCAAACTATCACTTCTGGGCATTCATTTTCTTCTGCTACTGAAA
TAACACCATCAGTGGCATTCACAGAAGTGCCATCTTTATTTCCTTCTAAAA
AGAGTGCAAAAAGAACAATTTTATCCTCATCCTTGGAAGAATCCATTACC
CTATCAAGTAATTTGGATGTTAATTTATGTTTGGATAAGACTTGTTTATCC
ATTGTCCCTTCACAAACTATCTCTTCAGACTTGATGAATTCTGATTTGACT
TCAAAAATGACTACTGATGAACTGTCAGTATCAGAAAACATTTTAAAACT
ATTGAAAATAAGACAATATGGCATAACTATGGGACCCACTGAGGTACTAA
ATCAAGAGAGCTTATTGGACATGGAAAAAAGTAAAGGATCTCATACACT
GTTCAAACTTCACCCAAGTGATAGTTCTCTGGATTTTGAGTTAAACTTACA
AATTTATCCGGATGTTACTTTAAAGACATATTCAGAAATTACACATGCAA
ATGACTTCAAAAATAATCTGCCACCATTGACAGGCTCAGTGCCTGATTTTT
CAGAAGTCACCACCAATGTTGCATTTTATACAGTTTCAGCAACTCCAGCA
CTTTCAATACAGACGTCTTCCTCCATGTCTGTAATTAGGCCAGATTGGCCA
TATTTTACAGATTATATGACCTCTCTTAAAAAAGAGGTCAAGACTTCTTCA
GAATGGTCCAAATGGGAACTTCAGCCTAGTGTGCAATATCAGGAATTTCC
CACTGCAAGCCGGCATCTTCCCTTCACTAGATCTCTTACTTTGTCTTCACT
GGAATCCATTCTGGCACCTCAACGGCTGATGATTTCTGATTTCAGTTGTGT
TCGTTATTATGGAGATTCTTATCTAGAATTTCAGAATGTGGCTTTAAATCC
ACAAAATAACATCTCCCTAGAATTTCAGACCTTCAGCTCCTATGGACTTCT

GCTGTATGTCAAGCAAGACTCAAATTTAGTAGATGGATTTTTTATTCAATT
GTTTATTGAAAATGGTACTTTAAAGTACCACTTTTACTGTCCTGGTGAAGC
AAAATTTAAAAGCATTAATACTACTGTTAGAGTGGACAACGGGCAAAAGT
ATACACTGCTTATCAGGCAAGAATTGGATCCATGTAACGCTGAGCTGACT
ATTTTAGGGAGGAATACACAAATATGCGAATCTATCAATCATGTACTCGG
AAAACCCCTGCCAAAATCAGGATCTGTCTTCATTGGTGGATTTCCAGACC
TTCATGGGAAAATCCAGATGCCTGTACCAGTCAAGAATTTTACTGGCTGC
ATAGAAGTTATAGAAATAAATAACTGGAGATCTTTCATTCCATCCAAGGC
AGTGAAAAACTATCACATTAACAATTGCAGATCCCAGGGATTTATGCTGT
CTCCAACAGCCTCCTTTGTTGATGCTTCTGATGTGACACAAGGGGTTGATA
CCATGTGGACTTCTGTCAGCCCCTCTGTTGCAGCACCCTCTGTGTGCCAGC
AGGATGTATGCCACAATGGAGGCACATGCCATGCCATCTTCCTCTCCAGT
GGCATAGTGTCATTCCAATGTGACTGTCCACTACATTTCACTGGCCGCTTC
TGTGAAAAAGATGCAGGTTTATTCTTTCCATCTTTCAATGGGAATTCCTAT
TTAGAACTGCCCTTTTTGAAGTTTGTCCTGGAGAAGGAACATAACAGAAC
TGTTACCATCTACTTGACTATAAAAACAAACAGTTTAAATGGAACTATTCT
TTACAGTAATGGGAATAATTGTGGAAAGCAGTTTCTTCATTTATTTCTTGT
GGAAGGAAGGCCATCAGTTAAATATGGGTGTGGAAATTCTCAAAATATTT
TGACTGTTTCTGCTAATTACAGCATTAACACAAATGCATTCACCCCTATCA
CAATACGCTACACAACGCCTGTTGGCAGCCCTGGAGTTGTTTGTATGATT
GAAATGACTGCAGATGGAAAACCTCCAGTACAGAAGAAAGACACAGAGA
TTTCCCATGCCTCTCAGGCATATTTTGAATCAATGTTCCTTGGCCATATTC
CTGCAAATGTTCAAATTCATAAGAAAGCAGGTCCTGTTTATGGGTTCAGG
GGCTGCATTCTAGACCTTCAAGTAAACAACAAAGAATTCTTCATCATCGA
TGAAGCACGACATGGAAAGAATATTGAGAACTGCCACGTCCCTTGGTGTG
CTCATCATCTGTGCCGCAACAATGGCACCTGCATCAGTGATAATGAAAAT
CTGTTTTGTGAGTGTCCAAGGCTGTATTCAGGCAAGCTGTGCCAGTTTGCA
AGTTGTGAAAACAACCCATGTGGAAATGGTGCCACCTGTGTTCCAAAATC
CGGAACAGATATTGTCTGCCTCTGCCCATATGGGAGGTCTGGACCCCTCT
GCACTGATGCTATTAATATTACTCAGCCAAGGTTCAGTGGCACAGATGCC
TTTGGATACACCTCATTCCTGGCTTATTCACGGATCTCAGACATCAGCTTC
CATTATGAATTCCACCTGAAGTTTCAGCTGGCAAACAACCACTCAGCACT
GCAAAATAACTTGATATTTTTTACTGGACAGAAAGGCCATGGGTTGAATG
GCGATGACTTCCTGGCTGTGGGCCTGCTCAATGGCAGTGTGGTTTATAGTT
ATAACCTGGGGTCTGGCATAGCAAGCATCAGGAGCGAGCCCCTCAATCTG
AGCCTTGGAGTCCACACTGTTCATCTGGGCAAGTTCTTCCAAGAGGGCTG
GCTGAAGGTAGATGATCATAAAAATAAATCCATTATCGCCCCAGGAAGAC
TGGTTGGTCTCAATGTCTTCAGTCAGTTTTATGTAGGTGGCTACAGTGAAT
ACACTCCAGATCTCTTACCAAATGGAGCAGATTTTAAAAATGGTTTTCAA
GGCTGTATTTTCACTCTTCAAGTTCGCACTGAGAAGGATGGCCATTTCAGA
GGACTGGGAAATCCTGAGGGCCACCCAAATGCTGGACGCAGTGTTGGCC
AGTGTCATGCTTCTCCCTGCAGTTTAATGAAATGTGGCAATGGTGGGACA
TGCATAGAGAGTGGAACTAGTGTTTACTGCAATTGTACCACTGGGTGGAA
AGGATCATTCTGCACAGAGACAGTTTCTACCTGTGATCCTGAACATGACC
CTCCACACCACTGTAGCAGAGGAGCAACCTGCATTTCATTACCTCATGGA

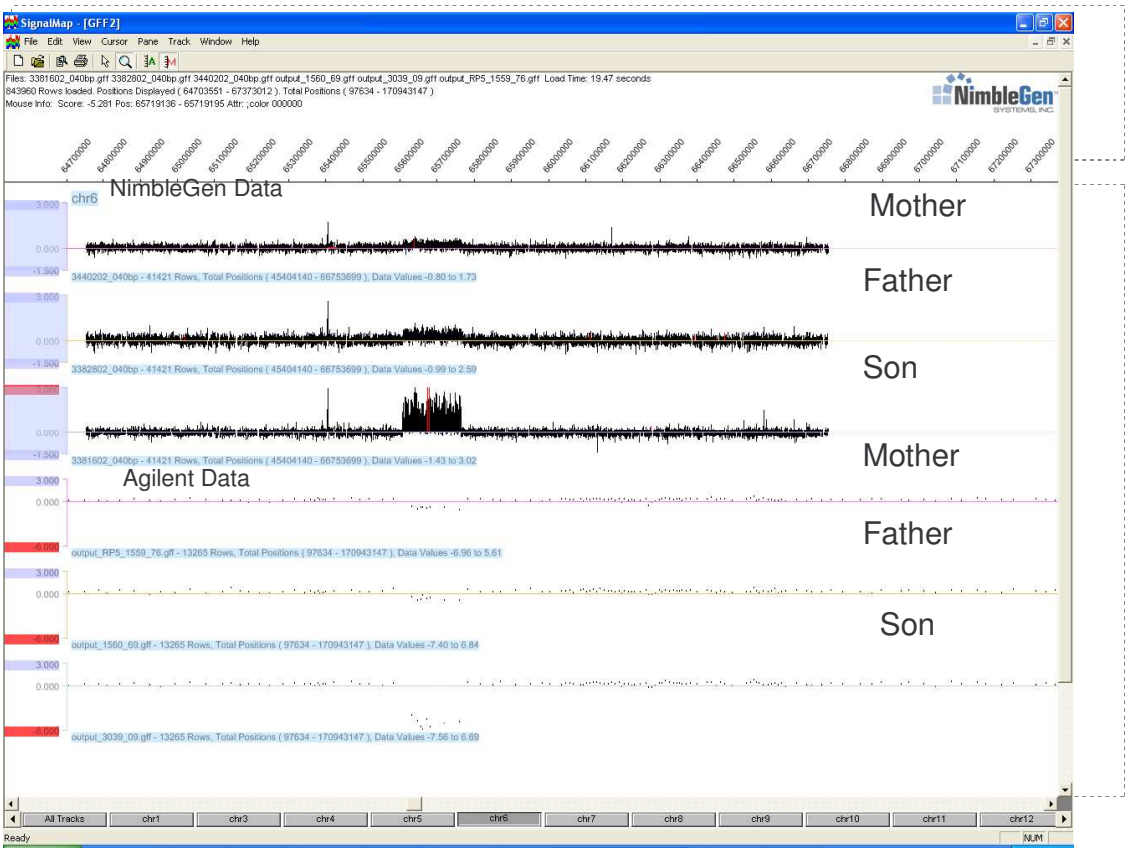
TACACCTGTTTCTGTCCTCTAGGAACCACTGGAATCTACTGTGAACAAGCT
TTATCCATAAGTGATCCATCTTTCAGAAGCAATGAGTTATCCTGGATGTCT
TTTGCTTCCTTTCATGTTCGAAAAAAGACACATATTCAATTGCAGTTTCAG
CCTCTCGCTGCAGATGGTATCCTATTTTATGCTGCACAACACTTAAAAGCC
CAATCAGGTGATTTTTTATGCATCTCTTTAGTAAATAGTTCCGTTCAACTT
CGCTACAACCTTGGCGACAGAACTATCATTCTAGAAACTCTCCAAAAAGT
AACTATAAACGGAAGTACTTGGCATATAATAAAAGCAGGAAGAGTTGGT
GCAGAAGGCTACCTGGATCTAGATGGGATAAATGTAACAGAAAAGGCCT
CCACTAAAATGAGTTCTCTGGACACAAATACAGACTTCTATATTGGAGGA
GTATCTTCTTTAAATCTTGTAAATCCCATGGCAATAGAAAATGAACCTGTA
GGTTTTCAAGGCTGTATCCGACAAGTTATCATAAATAATCAAGAATTACA
ATTAACTGAATTTGGAGCAAAAGGTGGCTCAAATGTAGGTGACTGTGATG
GGACAGCCTGTGGGTACAACACATGCAGAAATGGAGGTGAATGTACAGT
AAATGGCACAACTTTTTCTTGCAGATGTTTGCCAGATTGGGCTGGAAATA
CATGTAACCAGTCTGTGTCCTGTTTGAATAATCTTTGCCTCCACCAATCTT
TATGTATACCTGACCAATCATTTTCTTACAGTTGCCTGTGTACTTTGGGTT
GGGTGGGAAGGTATTGTGAAAACAAAACTTCATTTTCAACTGCAAAATTT
ATGGGTAATTCTTACATTAAATACATTGATCCAAATTATAGAATGAGAAA
CCTCCAGTTCACTACTATATCCTTAAATTTCAGTACCACTAAAACAGAAG
GTCTAATTGTATGGATGGGAATAGCTCAAAATGAAGAAAATGATTTTCTG
GCAATTGGTCTCCATAATCAGACCTTGAAAATAGCAGTTAACTTGGGAGA
AAGAATCTCTGTGCCTATGAGCTATAACAATGGCACATTCTGTTGTAATA
AATGGCACCATGTAGTTGTAATTCAAAATCAGACTCTTATCAAGGCCTAC
ATAAATAACAGTCTAATTCTTTCCGAGGATATTGATCCACATAAAAACTTT
GTGGCTCTAAACTATGATGGCATTTGTTATCTAGGGGGCTTTGAATATGGT
AGAAAGGTAAATATCGTTACTCAAGAGATTTTTAAAACCAATTTTGTTGG
CAAAATTAAAGATGTTGTATTTTTTCAGGAACCAAAAAACATTGAACTAA
TTAAATTAGAAGGATACAATGTTTATGATGGAGATGAACAAAATGAGGTT
ACATAAgttaacactagagattttagtacacactatacataaatgcagttattttgatagttatttctttgatacattgcttac
ggggaatcaactgtttactatattacctgaaatagtctaaatgctaacatatctttttgataagattgtaaaatgtcactgaaggtt
ctgattgtttttcctacatctaatttatctgtatatattttgattcatgttttaactccattagttcagtgcttattcacagaatgtgcttat
tcactttgcttattcacttttacctatccctaatgcttacattttaaaatcaacgtgtaaacacaattttaaaaatcaatgtgtaagc
tgaactttaaaatgtttaaaattcagttgggaatgaattagggataggtacaaataagaaaaaaatcaagttattaatcaaagg
aaaaataaaaattatacaaagctaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

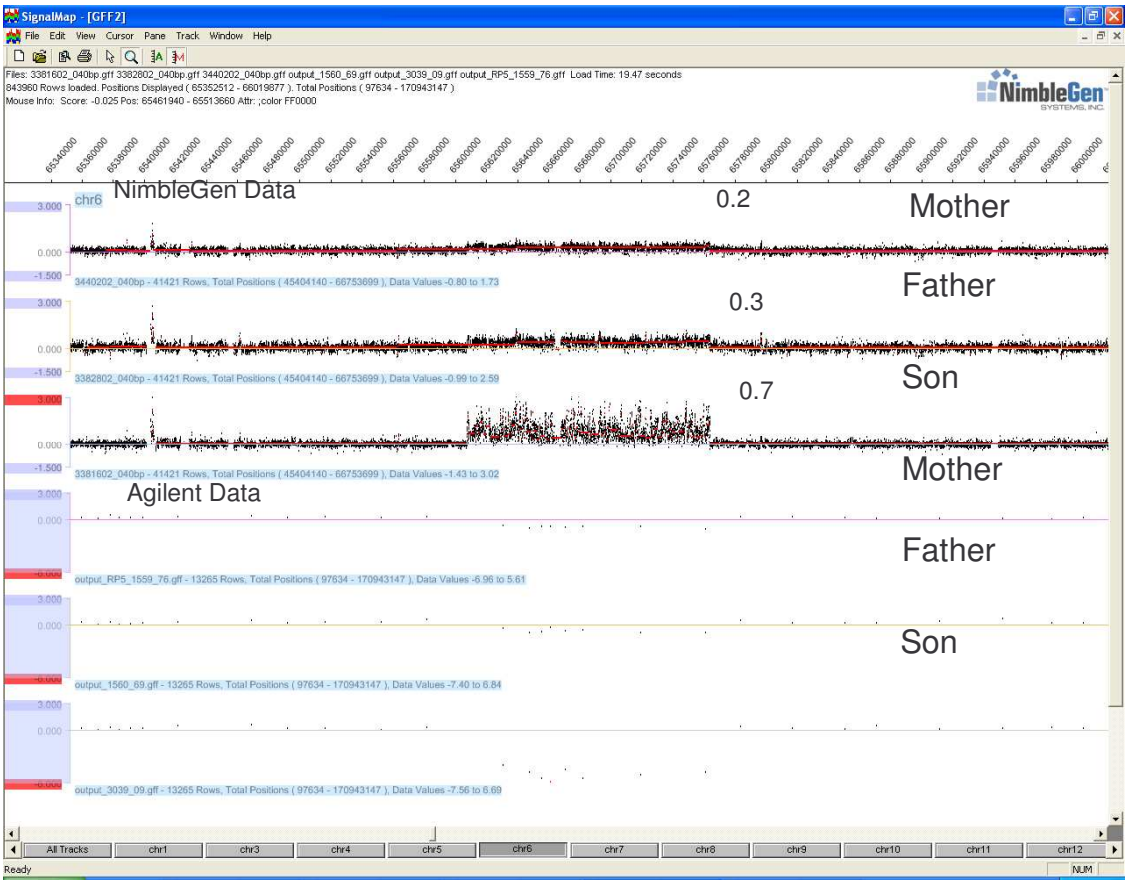
Alternative colours represent exons, 43 in total. Underlined and lower cases represent the 5'and 3' UTRs.

Protein Sequence of SPAM

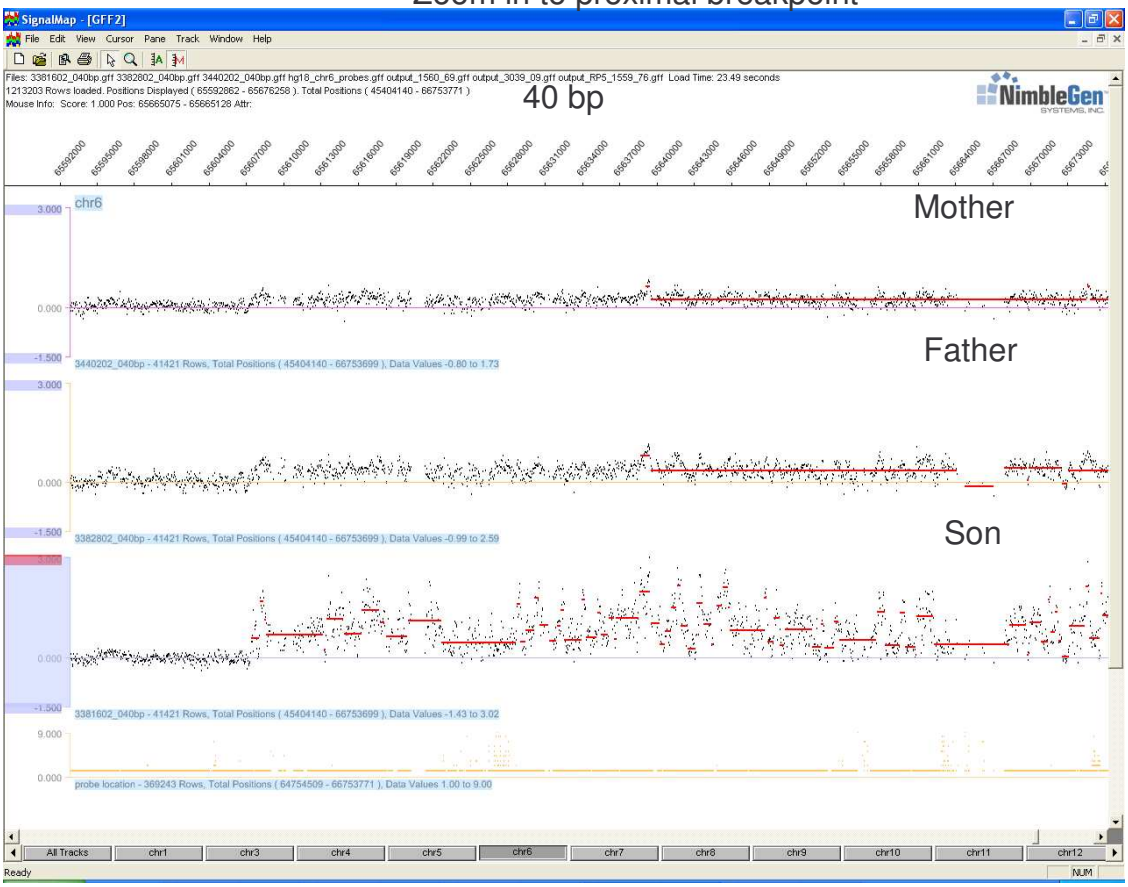
```
  1 MTDKSIVILS LMVFHSSFIN GKTCRRQLVE EWHPQPSSYV VNWTLTENIC
 51 LDFYRDCWFL GVNTKIDTSG NQAVPQICPL QIQLGDILVI SSEPSLQFPE
101 INLMNVSETS FVGCVQNTTT EDQLLFGCRL KGMHTVNSKW LSVGTHYFIT
151 VMASGPSPCP LGLRLNVTVK QQFCQESLSS EFCSGHGKCL SEAWSKTYSC
201 HCQPPFSGKY CQELDACSFK PCKNNGSCIN KRENWDEQAY ECVCHPPFTG
251 KNCSEIIGQC QPHVCFHGNC SNITSNSFIC ECDEQFSGPF CEVSAKPCVS
301 LLFWKRGICP NSSSAYTYEC PKGSSSQNGE TDVSEFSLVP CQNGTDCIKI
351 SNDVMCICSP IFTDLLCKSI QTSCESFPLR NNATCKKCEK DYPCSCISGF
401 TEKNCEKAID HCKLLSINCL NEEWCFNIIG RFKYVCIPGC TKNPCWFLKN
451 VYLIHQHLCY CGVTFHGICQ DKGPAQFEYV WQLGFAGSEG EKCQGVIDAY
501 FFLAANCTED ATYVNDPEDN NSSCWFPHEG TKEICANGCS CLSEEDSQEY
551 RYLCFLRWAG NMYLENTTDD QENECQHEAV CKDEINRPRC SCSLSYIGRL
601 CVVNVDYCLG NHSISVHGLC LALSHNCNCS GLQRYERNIC EIDTEDCKSA
651 SRKNGTTSTH LRGYFFRKCV PGFKGTQCEI DIDECASHPC KNGATCIDQP
701 GNYFCQCVPP FKVVDGFSCL CNPGYVGIRC EQDIDDCILN ACEHNSTCKD
751 LHLSYQCVCL SDWEGNFCEQ ESNECKMNPC KNNSTCTDLY KSYRCECTSG
801 WTGQNCSEEI NECDSDPCMN GGLCHESTIP GQFVCLCPPL YTGQFCHQRY
851 NLCDLLHNPC RNNSTCLALV DANQHCICRE EFEGKNCEID VKDCLFLSCQ
901 DYGDCEDMVN NFRCICRPGF SGSLCEIEIN ECSSEPCKNN GTCVDLTNRF
951 FCNCEPEYHG PFCELDVNKC KISPCLDEEN CVYRTDGYNC LCAPGYTGIN
1001 CEINLDECLS EPCLHDGVCI DGINHYTCDC KSGFFGTHCE TNANDCLSNP
1051 CLHGRYTELI NEYPCSCDAD GTSTQCKIKI NDCTSIPCMN EGFCQKSAHG
1101 FTCICPRGYT GAYCEKSIDN CAEPELNSVI CLNGGICVDG PGHTFDCRCL
1151 PGFSGQFCEI NINECSSSPC LHGADCEDHI NGHVCKCQPG WSGHHCENEL
1201 ECIPNSCVHE LCMENEPGST CLCTPGFMTC SIGLLCGDEI RRITCLTPIF
1251 QRTDPISTQT YTIPPSETLV SSFPSIKATR IPAIMDTYPV DQGPKQTGIV
1301 KHDILPTTGL ATLRISTPLE SYLLQELIVT RELSAKHSLL SSADVSSSRF
1351 LNFGIRDPAQ IVQDKTSVSH MPIRTSAATL GFFFPDRRAR TPFIMSSLMS
1401 DFIFPTQSLL FENCQTVALS ATPTTSVIRS IPGADIELNR QSLLSRGFLL
1451 IAASISATPV VSRGAQEDIE EYSADSLISR REHWRLLSPS MSPIFPAKVI
1501 ISKQVTILNS SALHRFSTKA FNPSEYQAIT EASSNQRLTN IKSQAADSLR
1551 ELSQTCATCS MTEIKSSREF SDQVLHSKQS HFYETFWMNS AILASWYALM
1601 GAQTITSGHS FSSATEITPS VAFTEVPSLF PSKKSAKRTI LSSSLEESIT
1651 LSSNLDVNLC LDKTCLSIVP SQTISSDLMN SDLTSKMTTD ELSVSENILK
1701 LLKIRQYGIT MGPTEVLNQE SLLDMEKSKG SHTLFKLHPS DSSLDFELNL
1751 QIYPDVTLKT YSEITHANDF KNNLPPLTGS VPDFSEVTTN VAFYTVSATP
1801 ALSIQTSSSM SVIRPDWPYF TDYMTSLKKE VKTSSEWSKW ELQPSVQYQE
1851 FPTASRHLPF TRSLTLSSLE SILAPQRLMI SDFSCVRYYG DSYLEFQNVA
1901 LNPQNNISLE FQTFSSYGLL LHVKQDSNLV DGFFIQLFIE NGTLKYHFYC
1951 PGEAKFKSIN TTVRVDNGQK YTLLIRQELD PCNAELTILG RNTQICESIN
2001 HVLGKPLPKS GSVFIGGFPD LHGKIQMPVP VKNFTGCIEV IEINNWRSFI
2051 PSKAVKNYHI NNCRSQGFML SPTASFVDAS DVTQGVDTMW TSVSPSVAAP
2101 SVCQQDVCHN GGTCHAIFLS SGIVSFQCDC PLHFTGRFCE KDAGLFFPSF
2151 NGNSYLELPF LKFVLEKEHN RTVTIYLTIK TNSLNGTILY SNGNNCGKQF
2201 LHLFLVEGRP SVKYGCGNSQ NILTVSANYS INTNAFTPIT IRYTTPVGSP
2251 GVVCMIEMTA DGKPPVQKKD TEISHASQAY FESMFLGHIP ANVQIHKKAG
2301 PVYGFRGCIL DLQVNNKEFF IIDEARHGKN IENCHVPWCA HHLCRNNGTC
2351 ISDNENLFCE CPRLYSGKLC QFASCENNPC GNGATCVPKS GTDIVCLCPY
2401 GRSGPLCTDA INITQPRFSG TDAFGYTSFL AYSRISDISF RYEFHLKFQL
2451 ANNHSALQNN LIFFTEQKGH GLNGDDFLAV GLLNGSVVYS YNLGSGIASI
2501 RSEPLNLSLG VHTVHLGKFF QEGWLKVDDH KNKSIIAPGR LVGLNVFSQF
2551 YVGGYSEYTP DLLPNGADFK NGFQGCIFTL QVRTEKDGHF RGLGNPEGHP
2601 NAGRSVGQCH ASPCSLMKCG NGGTCIESGT SVYCNCTTGW KGSFCTETVS
2651 TCDPEHDPPH HCSRGATCIS LPHGYTCFCP LGTTGIYCEQ ALILIVILEK
2701 PKPAERKVKK EALSISDPSF RSNELSWMSF ASFHVRKKTH IQLQFQPLAA
2751 DGILFYAAQH LKAQSGDFLC ISLVNSSVQL RYNLGDRTII LETLQKVTIN
2801 GSTWHIIKAG RVGAEGYLDL DGINVTEKAS TKMSSLDTNT DFYIGGVSSL
2851 NLVNPMAIEN EPVGFQGCIR QVIINNQELQ LTEFGAKGGS NVGDCDGTAC
2901 GYNTCRNGGE CTVNGTTFSC RCLPDWAGNT CNQSVSCLNN LCLHQSLCIP
2951 DQSFSYSCLC TLGWVGRYCE NKTSFSTAKF MGNSYIKYID PNYRMRNLQF
3001 TTISLNFSTT KTEGLIVWMG IAQNEENDFL AIGLHNQTLK IAVNLGERIS
3051 VPMSYNNGTF CCNKWHHVVV IQNQTLIKAY INNSLILSED IDPHKNFVAL
3101 NYDGICYLGG FEYGRKVNIV TQEIFKTNFV GKIKDVVFFQ EPKNIELIKL
3151 EGYNVYDGDE QNEVT
```

Supplementary Agilent and Nimblegen Data on RP5 family

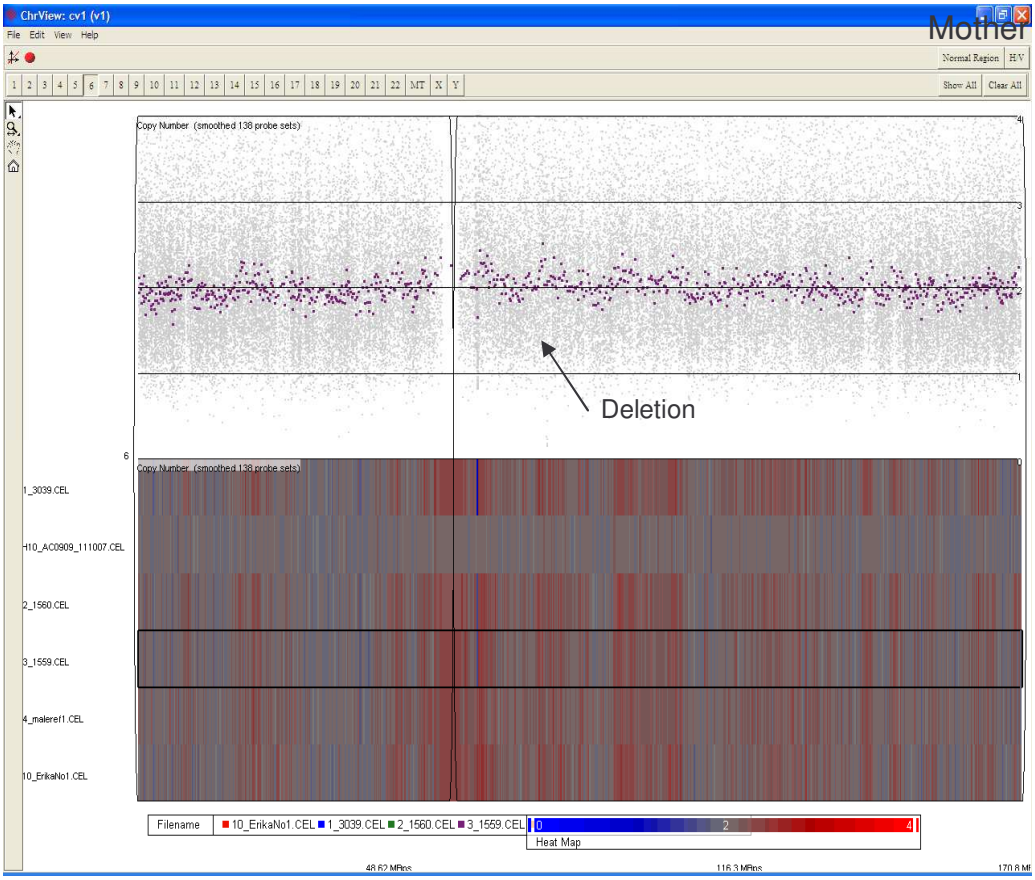



SignalMap - [GFF2]
File Edit View Cursor Pane Track Window Help
NimbleGen
NimbleGen Data
chr6
0.2
Mother
Father
0.3
Son
0.7
Mother
Agilent Data
Father
Son
All Tracks
chr1
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
Ready

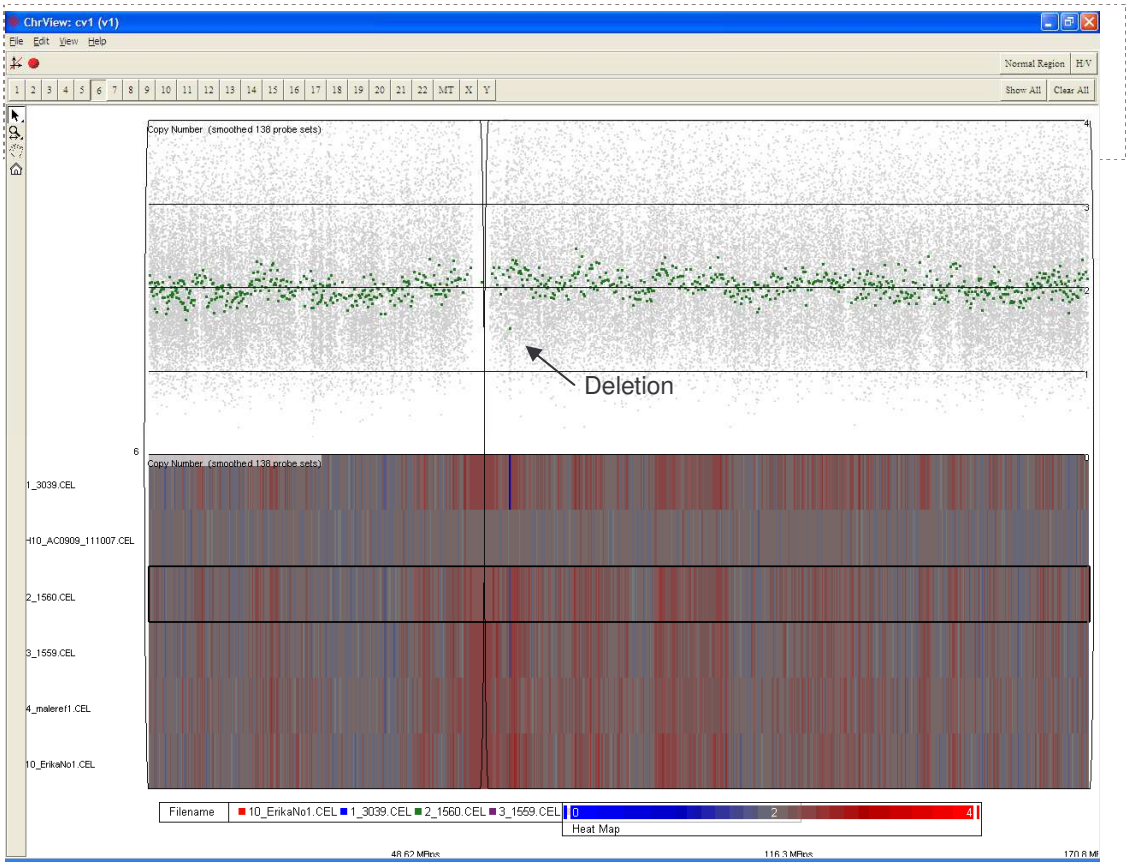
Zoom in to proximal breakpoint



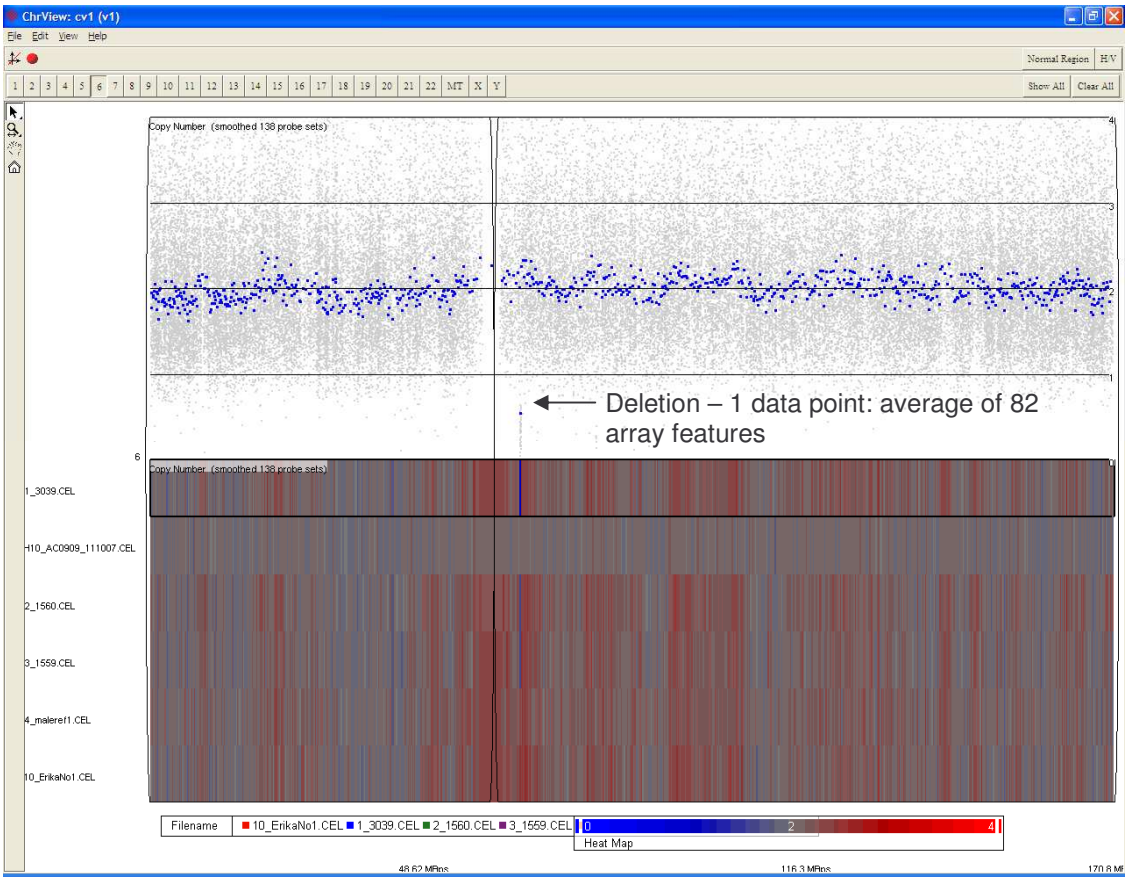
Affymetrix Data on RP5 family members



Father


ChrView: cv1 (v1)
File Edit View Help
Normal Region H/V
Show All Clear All
Copy Number: (smoothed 138 probe sets)
Deletion
Copy Number: (smoothed 138 probe sets)
1_3039.CEL
H10_AC0909_111007.CEL
2_1560.CEL
3_1559.CEL
4_maleref1.CEL
10_ErikaNo1.CEL
Filename
10_ErikaNo1.CEL 1_3039.CEL 2_1560.CEL 3_1559.CEL
Heat Map

Son


ChrView: cv1 (v1)
File Edit View Help
Normal Region
H/V
Show All
Clear All
Copy Number (smoothed 138 probe sets)
Deletion – 1 data point: average of 82 array features
Copy Number (smoothed 138 probe sets)
1_3039.CEL
H10_AC0909_111007.CEL
2_1560.CEL
3_1559.CEL
4_maleref1.CEL
10_ErikaNo1.CEL
Filename
10_ErikaNo1.CEL
1_3039.CEL
2_1560.CEL
3_1559.CEL
Heat Map

Primers

Primers for Mutation Screening of Candidate genes for RP25:

Gene	Exon No	Sequence (5`…… 3`)	Product size (bp)	$MgCl_2$	Temp
CYB5R4	Exon 1&2 F Exon 1&2 R	TGTTGACAGCCTCTTAAAACTG CACAAGGCATATTTTTGTTAAATTG	398	1.5	58.5
	Exon 3 F Exon 3 R	CCATGAAAGTGGCAGGTTAAG AAGGGCAAATTAAATGAACAGC	474	1.5	58.5
	Exon 4 F Exon 4 R	TGCAGCCTTTCTATCACACC TCGAACTTTTAAGGGAAAGTCAC	448	1.5	58.5
	Exon 5 F Exon 5 R	TGGTCTGGCAGTCTTCAGAG GGCAGGAGAATTGCTTGAAC	437	1.5	63.2
	Exon 6 F Exon 6 R	GGCCTTTTGGCTAAGATCAAG TGGAACAGCTCATTCTCCAAC	393	1.5	60.8
	Exon 7 F Exon 7 R	GCTGTATAATGAAGAATGTAACATGG GCCACTCAGTGTGTAACTTTGG	426	1.5	58.5
	Exon 8 F Exon 8 R	TGAAGGAGAAGAACTGGGTTG GTCACACCGCCACTTGCTA	452	1.5	60.8
	Exon 9 F Exon 9 R	TGTTTTGGTCTTTCAGAACACG CACATCTTTGTTTTGCAATGTG	422	2.5	58.5
	Exon 10F Exon 10R	TCCCACTAATAATCAGTACCATTCA CAACACACATTGTCCCTTTCA	399	2.5	60.8
	Exon 11F Exon 11R	GGCATTTTAGTGAAATTGAAGC CACGTTGCAGTCACAAACCT	309	2.5	60.8
	Exon 12F Exon 12R	TGAAAATTGTTTCGGCACAC CCGACTCGAGTTTCCAGATT	572	2.5	58.5
	Exon 13+14F Exon 13+14R	TTAGGGGCCTAATGAGTATCC TGCCTAGATCTTGCCAGACA	940	2.5	58.5
	Exon 15F Exon 15R	TTAAATTGTTCCTTTAGGGGAGA TAATCCTGGCCTCCAACTCA	441	2.5	63.2
	Exon 16F Exon 16R	TGAGCTAGTTGAAAATTTGAGCA TGTGGTACAAAAGAAAGCCAAG	456	2.5	63.2
ME1	Exon 1 F Exon 1 R	GATGATCTCCTGCCATGACTC CTTCTTCCTTCTGCCCACAC	492	1.5	58.5
	Exon 2 F Exon 2 R	GGCCATGCTTCCAGTGAC TGAGCACGCAGTTTAGAAAGC	476	3.5	58.5
	Exon 3 F Exon 3 R	GGCCAAGACTCTGTCGTAAAG TGTATACATGTTTCAGAACATCCTC	479	1.5	58.5
	Exon 4 F Exon 4 R	GGAAATAATGTCCACCAGTTTGA TTCACTAGAAAAGCTAATGAATGC	333	3.5	58.5
	Exon 5 F Exon 5 R	TTCTTCACCCTATTCTTCCACAA TGGGAAAAGGAGAACAGGAG	104	1.5	57.0
	Exon 6 F Exon 6 R	TTTCTATTTTTGTCCCTTTTATACAGA CATTTGTCCATTAACAGTTATGAGG	600	DMSO	58.5
	Exon 7 F Exon 7 R	TGTTTTCCTATTTCTATGGGCTAA AGTATGTGACATCAGCAAGCAT	338	1.5	57.0

	Exon 8 F Exon 8 R	CGTATGCAAGCATGGGATAA CGGCAACATTTGATAATACCAC	483	1.5	57.0
	Exon 9 F Exon 9 R	TGGCTTAACCTAAATGTGACAAAA TGACATGATCATTTGTATTTAATGG	375	1.5	58.5
	Exon 10 F Exon 10 R	CGAATGAATGAAATAACATTCTGA TCTACAGGGTACTGTCGCTTGA	414	1.5	58.5
	Exon 11 F Exon 11 R	TTGTTGGCATTAACAAACTTGC TCCCTCTGATTTGTGTGCTTT	465	1.5	57.0
	Exon 12 F Exon 12 R	CATCACAAAAGCTTGTCCAGA AACATGCTACCCACCCCTTA	492	3.5	60.8
	Exon 13 F Exon 13 R	TGTTGATAGAAGTGGGGATGAA TCAGTGTGCCTCATTTCTCTG	399	1.5	58.5
	Exon 14 F Exon 14 R	CACACTTTCCTTAGTTGCTGATG ACCAGAAGGCAAAGTGAAGC	597	1.5	60.8
MTO1	Exon 1 F Exon 1 R	ATCACCCATCTCCCTCACC CTGATCCTTGCGAGACACTG	559	1.5	60.8
	Exon 2+3 F Exon 2+3 R	GACACTTTTGCATCCCATTG GCAACATATCTCCACACTTTGC	662	1.5	60.8
	Exon 4 F Exon 4 R	GCTCACGCCTGTAAGTAATCC TACTCAAGGGGGTTGAGGTG	910	1.5	60.8
	Exon 5+6 F Exon 5+6 R	GGGATGTTTATTTTTGGCAAG GAGCAAAACTCCGTCTCAAAAC	736	1.5	60.8
	Exon 7 F Exon 7 R	TGCCCTTCAGTATAAGGTCAGC TGAGGTCGGGAGTTTGAGAC	361	1.5	60.8
	Exon 8 F Exon 8 R	GCATTTAAGGGTAATGCTTGC AAAAGCCATCTCCAAACACC	446	1.5	65.0
	Exon 9 F Exon 9 R	TTAGGGGCAATACTTGCTTTC TGCCCAGGCTAGTCCTAAAC	434	1.5	57.0
	Exon 10 F Exon 10 R	GGCAAGAGGATCACTTGAGG TGAAAAGAAGCCAAAAACAAGC	433	1.5	57.0
	Exon 11 F Exon 11 R	CTACAGTTGATGGGCGACAG AGGTGTTGCAATGCTCTTAGC	433	1.5	57.0
	Exon 12 F Exon 12 R	AAGGGAACACCTACCCAACC AGAGATGGGGTCTTGCTGTG	467	1.5	57.0
	Exon 13 F Exon 13 R	CAGATTGTTATTTGGAGGAAAATAG AAGGGCTTTGATATTGCTTCC	670	1.5	57.0
DPPA5	Exon 1+2 F Exon 1+2 R	ACCAGGTGCGAGGCTTTATC CCGCCTTTCACCATAGTCTC	585	1.5	57.0
	Exon 3 F Exon 3 R	CCCCCTTCAGTTTCTTGGAG GTTCAGTGCTGCATCTCAGG	492	2.5	58.5
SNX14	Exon 1 F Exon 1 R	CGAGACGCTAGAGGCTTTG CTCCGCACGGTTAAGCAG	491	2.5	58.5
	Exon 2 F Exon 2 R	TGTAGGTGCTTAATGAATGCTTG GCAGGCTTTCAAATAATAAACAG	290	1.5	57.0
	Exon 3+4 F Exon 3+4 R	TGATGGTTTAGTTTAGCCAGTATTATG TGCCTCAGAACTGATTTAAGTGAC	546	2.5	65.2
	Exon 5 F Exon 5 R	CTCCCAAAGTGCTGGGATTA ACAGTACTGCCCCCAAAACA	369	2.5	60.8
	Exon 6 F Exon 6 R	TCGTGTATGTCTGCAAGGAAA AACACAAATGATAAATGCTTGAGG	538	1.5	57.0
	Exon 7 F Exon 7 R	AGGGCCTAACCAAGACCTAC AGTTCTGGAGATGGGGAATG	436	1.5	58.5

	Exon 8/9/10 F Exon 8/9/10 R	GCCCATTTTCACAAAGGTTAAG AGTGAAATTCCGCCTCAAAC	807	1.5	58.5
	Exon 11 F Exon 11 R	AAAACTGGTTGATTGCATTTCTC CCAACCAAGCTCTAGTTCTGAC	536	1.5	60.8
	Exon 12 F Exon 12 R	TGGCTAGCTGCAGTACATGG TGGTCTTGAATGCCTGAACTC	548	1.5	60.8
	Exon 13 F Exon 13 R	TTGCAATATTTTCATCCTTACCC TTCAGGCATTTTCAGATCACC	398	2.5	60.8
	Exon 14 F Exon 14 R	CCGTGTATGAAAACTCCAAAAG AATGAAAGAATGAGAAATACAAAGG	451	2.5	62.0
	Exon 15 F Exon 15 R	CCCCGGCCTATTATATTTTC TGTTCTGGACACTTCTCTTTTGG	574	1.5	58.5
	Exon 16 F Exon 16 R	TTCTGAAGCTTTTAAATTTTGTTTT GGCAGTTAATACCACAGTGTTCC	247	1.5	62.0
	Exon 17 F Exon 17 R	CTTGAAGCTCTCTGGTTTTGTG TTGACAAAAATTCATTCTAAAATTCAC	429	1.5	58.5
	Exon 18 F Exon 18 R	ATGTGGAAGAGCCAGGATTC TGACAAATGGCTAGAAGTGACTG	537	1.5	60.8
	Exon 19+20 F Exon 19+20 R	TTGAAAATTTCCATGGAGTGTG ACCGTTCAACTGATTTCCAC	592	1.5	58.5
	Exon 21 F Exon 21 R	CAATTCGGGGCATAAATGG AATATGATCAAGGCCAATCG	446	1.5	57.0
	Exon 22 F Exon 22 R	GATCTTGAGTATTGCCCCATC GGAGAGAGCGAGGTTCAGTG	585	1.5	63.0
	Exon 23 F Exon 23 R	AATATTGATAAATTTTGCTGGATTATG CACAATACCCTCTGTATACCGTTC	358	1.5	57.0
	Exon 24 F Exon 24 R	GGTGATTTTGCTTTATGATCCAG AATGTTTAAAATTATATTGCCTTGG	300	3.5	63.0
	Exon 25 F Exon 25 R	CCCTGCCTTCCCACTTAGAG AACCACAGAGAACAGAGACAAGG	278	3.5	63.0
	Exon 26 F Exon 26 R	CAGAGCTGTCATTAAACGTTGC AGAGGCAACCACACAATCAC	807	2.5	62.0
SENP6	Exon 1 F Exon 1 R	CCTGGCCTGCCTTTGTATAG GGCTCGTAGCTCAATCCAC	468	1.5	58.5
	Exon 2 F Exon 2 R	TGAGCTTCTTTGTGTAGCAAGC TTTCTTTCCTTAAAGCGACTGTG	470	1.5	60.8
	Exon 3 F Exon 3 R	TTGGAACATAGCCACATGTATTG CAATTGTGTAACATGGGAGGTG	490	1.5	60.8
	Exon 4 F Exon 4 R	TTCTTAAATCCATGCCAATGC TGCAATGTTTTATTCAGAAACAAG	500	1.5	57.0
	Exon 5 F Exon 5 R	GGAAATGGAAAACTCCCAAAAG TCTCAGTCAAACACTTTCAGGTC	566	1.5	57.0
	Exon 6+ 7 F Exon 6+ 7 R	TCCTTGATATTGTGTGTATTTGAAG CTGTGCAACTGTTCCAGGTATC	673	1.5	57.0
	Exon 8 F Exon 8 R	CAGGAAGGGCTTCTGTTACC TATGTTGCAGGGAAGCACTG	599	1.5	60.8
	Exon 9 F Exon 9 R	TGATGGTAATGATGGGGAAG ACCTGGCCGAAGTCAGAATG	737	1.5	63.2
	Exon 10 F Exon 10 R	CCAAGTCTTAATGCTTTCTTCAG TGGCTTAATGTTCTTCCAAATAATC	606	1.5	65.2
	Exon 11 F Exon 11 R	ACATTTCTTATATTGGGTCTTTGC CGAAGCTGGCATGATAATAAAC	486	1.5	65.2

	Exon 12+13 F Exon 12+13 R	TGATGGAAGGGATGTGACAA TTTCACAAGGTGGCTGTGAC	975	2.5	60.8
	Exon 14 F Exon 14 R	AATGTGATTCTGGTTCCTTCCAA GGAACCAAAAATGCTTCTGAG	670	1.5	68.0
	Exon 15+16 F Exon 15+16 R	AAAGTGGCCCCTAGTTTTGC AAAATGTGCAAAGACAGAAACAG	685	2.5	63.2
	Exon 17 F Exon 17 R	GGCTGAAGCAGGAGAATCAC AACCCATGAAGAATGAAAATGG	468	2.5	63.2
	Exon 18 F Exon 18 R	TGTCAAACCATTTTATGCAACTG CTGCATTAATTCTTGTAAAACAGG	378	2.5	60.8
	Exon 19 F Exon 19 R	CTCCCAAAGTGCTGGGATTA ATTGGGTCTCGCTATGTTGC	814	1.5	67.0
	Exon 20 F Exon 20 R	AAGGCAGATTTTGTGTTCTTGAC TTATGTTGCCCAGGCTGAAC	386	2.5	60.8
	Exon 21 F Exon 21 R	TGTTGTAGCATTGGATAATTGG AACTGGCAAATGGAGGACTG	471	2.5	60.8
	Exon 22+23 F Exon 22+23 R	CATACAAAGGGGATTTTGAAGG TTATGCCTTGAAGGTTCTTCC	600	2.5	63.2
	Exon 24 F Exon 24 R	AATTTGGTCCACTGGCAAAC TGAAAGGCTGTTACTTTGTTCC	961	2.5	60.8

Primers for 5` and 3` RACE PCR:

5` OR 3` RACE	EXON	SEQUENCE (5`……3`)	GC CONTENT	TM (MELTING TEMPERATURE)
3`	UTR F	CTCCACAGGTCAGGGCATCTGTAC	58.0%	76°C
3`	UTR R	GTACAGATGCCCTGACCTGTGGAG	58.0%	76°C
3`	EXON 2R	CTGTAAATGGTGGGTGACAGACAC	50.0%	84°C
3`	EXON 14F	TGACTGCAGGAAAGCCTGCTTATAC	66.0%	84°C
3`	EXON 14F	GGA GCCACCTGCA TTGACCAACC TGG	61.5%	84°C
3`	EXON 16F	GCAG TTCCAGCCCTTGCCTGCATG G	65.0%	82°C
3`	EXON 17F1	CATCTGGATG GACTGGACAGAACTGTAG	50.0%	84°C
3`	EXON 17F2	GTCTGTGCC CACCCCTTTA TACTGGAC	55.5%	84°C
3`	EXON 20R1	CCAGGGGCACAGAGGCAGTTGTATCC	66.0%	84°C
3`	EXON 23F	GGAGGGATCTGTGTTGATGGGCCTGG	66.0%	84°C
3`	EXON 25F	CCCCCTTCTGAGACTTTGGTCAG	57.1%	82°C
3`	EXON 26F1	CCGCAGATGTTTCCTCTTCTCGATTC	64.0%	78°C
3`	EXON 26F2	GGAGATTGCTCAGCCCCTCGATG	55.0%	72°C
3`	EXON 32 F	CTTTCCATCTTTCAATGGGAATTCC	50.0%	70°C
3`	EXON 35F	GGGTTCAGGGGCTGCATTCTAGACC	60.0%	80°C
3`	EXON 43F	CAACCTTGGCGACAGAACTATC	50.0%	70°C
5`	EXON 2CR	GTCCCAGTGGACAAGGTGATGGACC	66.0%	84°C
5`	EXON 3R	CAGACATGTGGTTGACACTGGCC	56.0%	78°C
5`	EXON 4R	GGTTTTGCTGACACCTCACAGAATGGAC	56.0%	78°C
5`	EXON 4R1	CCATGCAGGCAAGGGCTGGAACTGC	65.0%	82°C
5`	EXON 3R	CAGACATGTGGTTGACACTGGCC	56.0%	78°C
5`	EXON 4R	GGTTTTGCTGACACCTCACAGAATGGAC	56.0%	78°C
5`	EXON 4R1	CCATGCAGGCAAGGGCTGGAACTGC	65.0%	82°C
5`	EXON 7R	GGTGACTCCACAGTAGCAGAGGTGTTG	50.0%	72°C
5`	EXON 8R	CATCAATAACCCCTTGGCACTTTTCGCC	66.0%	84°C
5`	EXON 9R	CCAGATACATGTTGCCAGCCCATCTGAG	56.0%	78°C
5`	EXON 10AR	GGGCCAGGCAGAGGCCATGCACTG	54.1%	74°C
5`	EXON 10BR	GGGACACACTTGCGGAAGAAATATCCC	56.5%	72°C
5’	EXON 11R	AGCTCATTCTCACAGTGGTGTCC	53.8%	80°C
5’	EXON 14R	GTGCTCACAGGCATTCAGGATGCAG	52.0%	78°C
5`	EXON 17R1	CTACAGTTCTGTCCAGTCCATCCAGATG	50.0%	84°C
5`	EXON 17R2	GTCCAGTATAAAGGGGTGGGCACAGAC	55.5%	84°C
5’	EXON 18F	CCTGCCAGGATTATGGTGACTGTG	54.0%	74°C
5’	EXON 18R	CACAGTCACCATAATCCTGGCAGG	54.1%	74°C
5`	EXON 20R1	CCAGGGGCACAGAGGCAGTTGTATCC	55.0%	84°C
5`	EXON 20R2	CAGAAGGGCCCATGGTACTCAGGTTCAC	57.1%	88°C
5`	EXON 22R	CATTGTGTGCTAGTCCCATCTGCATCAC	50.0%	84°C
5’	EXON 23R	CCAGGCCCATCAACACAGATCCCTCC	55.0%	84°C
5`	EXON 26R	GGAATTTCCCACTGCAAGCCGGC	50.0%	72°C
5`	EXON 32R	AGAAGGAACATAACAGAACTGTTACC	50.0%	72°C
5`	EXON 36R	GGGAGGTCTGGACCCCTCTGCACTG	68.0%	84°C
5’	EXON 41R	GACCCTCCACACCACTGTAGCAGAG	60.0%	80°C
5’	EXON 43R	GGGGGCTTTGAATATGGTAGAAAGG	55.0%	82°C

Primers for amplification of full length *RP25* gene

Exon No	Sequence (5'…… 3`)	$MgCl_2$	Temp-C
4 F 26 R	CCAATTCCCAGGAATCCTTAACCACAAC GAATCGAGAAGAGGAAACATCTGCGG	2.5	57°C
26 F 43R	CCGCAGATGTTTCCTCTTCTCGATTC CCCAACCCAAAGTACACAGGCAACTG	2.5	58.5°C

Primers for expression analysis of cDNA of EYS

Exon No	Sequence (5`…… 3`)	$MgCl_2$	Temp-C
4F 11R	CCAATTCCCAGGAATCCTTAACCACAAC CCAGATACATGTTGCCAGCCCATCTGAG	1.5	58.5°C

Primers for MLPA Analysis:

Exon No	5` Probe sequence	3` Probe sequence
12	TGAAGACTGCAAATCTGCGTCCTG	CAAAAATGGAACAACTAGTACACATTTAAGGGGA
13	TGGAGCCACCTGCATTGACCAACCT	GGTAATTACTTCTGCCAGTGTGTGCCT
15	TGTGAACAAGAATCCAATGAGTGTAAAATGAATCC	TTGCAAGAACAATTCCACCTGTACTGACCTTTAC
16	TACATCTGGATGGACTGGACAGAACTGTAGTGAA	GAAATAAATGAATGCGACTCTGATCCATGCATGA
17	TCTCATACTCTTTGCAGAATTTGAAGGTAAAAACTG	TGAAATTGATGTGAAAGACTGCCTCTTCCT
18	TGGGTTTTCTGGATCTCTGTGTGAAATTGAAATT	AATGAATGTTCCTCTGAACCTTGCAAAAATAATGG
19	TTGTGAACCTGAGTACCATGGGCCCTT	CTGTGAACTTGATGTAAATAAATGTAAAATCTCACCTT

Primers for SPAM gene:

Gene	Exon No	Sequence (5`…… 3`)	$MgCl_2$	Temp-C
EYS	1(nc)F 1(nc) R	AGCTCCAGATGGTGATTTGC TCCTTAAGCAAATGACAGAGAAAA	1.5	60.8
	2(nc)F 2(nc) R	GGCAGGATACTGGCTAATCTG GCCAAAAACGATAAAACCTGA	1.5	60.8
	3(nc)F 3(nc)R	CCATCCTCTGGATTATCATAAAAG CTTTGGGAAAGAAGGCCAAG	1.5	60.8
	4a F 4a R	GCTGCTGGTGACACTATCTTTG GAAGTCCAAGCAGATGTTTTCTG	1.5	60.8
	4b F 4b R	GCCTGATGGTTTTTCACAGC AAAATGGAGGCTGGCAATG	1.5	60.8
	4c F 4c R	GCAGTTCTGCCAGGAATCTC GTGCTGGGATTACAGGTGTG	1.5	60.8
	5 F 5 R	TTTCAATTTAAATGCATCATCG TGAAAAGCATGTGAACTGTTG	2.5	57.0
	6 F 6 R	TTTGCAAAAGTTACTGTAGAATTGC GACCGTTCTTGTTCGTCTGAG	1.5	57.0
	7 F 7 R	TGAGATGGGAGATGGTGTTG CAACAATTAACCCAAAACATGC	1.5	57.0
	8 F 8 R	GCTTTTTGGCTAAGATCACAGG TGGCTAAGATTAATAAGAGCATTTG	1.5	57.0
	9 F 9 R	GGCTTTTGAACATGGATATGAC TCTCTTGCACCAAGTAGATTTCC	1.5	57.0
	10 F 10 R	GGAACTTATTTTGTGGCAGATG TGATTCTTCAAAATTTTTACTTTCC	1.5	57.0
	11. F 11.R	CAAGCTTTGAACCCTTGTCC CTTCTTCCCTCCTTTTATTGTGC	1.5	57.0
	12 F 12 R	GCCAAGAATGGACACTTTAT CCAAAGAAGCAATCCTATTATTCAA	1.5	57.0
	13 F 13 R	TGGCATTCTTATCTAATAAATTTGG TTGGTTGGTCACTTTAGAAGC	1.5	57.0
	14 F 14 R	TTGGTTAAAAGTGAACTACATAAATG A TGCTTGAGTTTCTGTTTTCTAACC	1.5	57.0
	15 F 15 R	GAGATATCAAAATGGCCAGGAG ATGATTGCGACACCATCTTG	2.5	60.0
	16 F 16 R	TCACAGGAAATTAGGCAAACAA GGATTTTTCCAACCCATTTT	2.5	60.8
	17+18 F 17+18R	CTGGCATGTTTTTATGCACAC CATTTTGCTCAGGCACACAT	2.5	58.0
	19 F 19 R	AAAATTTTGCAAGGAGAATTGC ATTTTTGCCCTGTTTGCATC	2.5	58.0
	20 F 20 R	TGTGCTTTGTTTTTGTCATTCAC AACTGGCAGCATCTGTCATC	1.5	57.0
	21 F 21 R	TTTCACCTGAACTAGGAAAGAAAAG CAACAAGAGACAAAGAAAGAAAAG G	1.5	57.0

	22 F 22 R	GAGGAAGGAAATGTCAAACAAG TTGCAGAGTGCATTACTAGTGG	1.5	57.0
	23 F 23 R	GTTGGAGCTCTGAAAACACG TAGTATTGGTGGAGTGGATTGTC	1.5	58.0
	24 F 24 R	CCACGGATAAGAGCTGAGAC AGAGAAGGAGAGATGCGCTG	1.5	57.0
	25 F 25 R	TGTCTATGGAAATGCAAATGGA AAACAGGAGTCATAACCAATAATGC	1.5	58.0
	26 (a) F 26 (a) R	AGTGCCAAAGTGGTTCGTTT GGGGTCCTTGCTCTCCTATC	1.5	58.0
	26 (b) F 26 (b) R	GCACCGGTTCAGTACAAAAG AGTTTGTGAAGGGACAATGGA	1.5	58.0
	26 (c) F 26 (c) R	ATCCTCATCCTTGGAAGAATCC GACCATGACAGGCTCTTTCTTC	1.5	57.0
	27 F 27 R	TTGAAAGAGGCAGGAAAGAGAC AAGAGACATCCTGGTGGTGAGT	1.5	61.0
	28 F 28 R	GCTTTTCTCAACCATGTTCCTC GGGGATAGGGTCACCTTAAAAC	1.5	57.0
	29 F 29 R	GATTAATCTGCTTCTGGCTTTG TGGAAAAACAGACTGACATTGG	1.5	57.0
	30 F 30 R	CCCATGTTCATTGCAGCATTAT GAACGTAGGAATGTGAAGCAAA	1.5	58.5
	31 F 31 R	AGGGTCATGTTATGTGGCTCAG CAGCTGTTTCTTGTTTGTGC	1.5	58.5
	32 F 32 R	TTCAGTCTTTTCCTCTGTACTGG GCTCTGAAACATTTGCAGCAT	1.5	58.5
	33 F 33 R	ACTCCTACCAACCCCCTAAATC GTGGTGGTGCACATCTGTAGTC	2.5	58.5
	34 F 34 R	TTCTGAAAGCATTCCATGTCC TTTTCTGGTGCTTTGTTGAGAG	2.5	58.5
	35 F 35 R	GCCAACAATAGCAACCTCTTTT ACATGTGTGCATCATTTAGGT	2.5	58.5
	36 F 36 R	GCATATGTGTTCATGCATGTGT CTGCTTGGTGATCAGTCTCACT	1.5	58.5
	37 F 37 R	AGATGCATCAGCAAAACGTAAC GCATCTAGGCAAAGGTCTTTT	1.5	58.5
	38 F 38 R	GAATAACAACAGCCAGTTGCAC CTGTGAACTTCGTGGATGTAGG	1.5	58.5
	39 F 39 R	TAACAGACACCAGCAGAGAAGC TGTTCAAGTCTGAAAGCAATCC	2.5	58.5
	40 F 40 R	TTCTCTGCGCATTTCTGTATTC CTGTCCTCCCATCATGTAACAA	1.5	58.5
	41 F 41 R	GACAAGTTAGCATCAGGGCAAT GAAAAAGAGGACAGTGGATTTG	1.5	57.0
	42 F 42 R	CTCACCTACAAGCAACTCTTGG TACGCATACACTTGCAGTGACA	1.5	57.0
	43 (a) F 43 (a) R	CTTATCCAACTTGGCCAGAAAC TCAAACAGGACACAGACTGGTT	1.5	57.0
	43 (b) F 43 (b) R	GGGTACAACACATGCAGAAATG GTGGATCAATATCCTCGGAAAG	1.5	57.0
	43 (c) F 43 (c) R	GGTCTCCATAATCAGACCTTGA GATTCCCCGTAAGCAATGTATC	3.5	57.0

Primers for Sequence Homology of EYS across species

Species	Gene and exon	SEQUENCE (5` □……..3`)	Size (bp)	Mg	Temp
Human	*EYS* exon 14	TGCTTATGCAATCCAGGCTA AGATGCAGGTCTTTGCAGGT	104	1.5	58.5
	EYS exon 26	AAGCACGACATTCTCCCAAC CCCCCTAGAGACAACTGGAG	543	1.5	58.5
	EYS exon 43	ACAACCTTGGCGACAGAACT CCACCCAACCCAAAGTACAC	552	1.5	60.8
	GAPDH exon 4	TGCATTCGCCCTCTTAATG ATCAGCCAGGTGGCTCCTC	322	1.5	60.8

nature genetics

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

EYS, encoding an ortholog of *Drosophila* spacemaker, is mutated in autosomal recessive retinitis pigmentosa

Mai M Abd El-Aziz[1,8], Isabel Barragan[2,3,8], Ciara A O'Driscoll[1,8], Leo Goodstadt[4], Elena Prigmore[5], Salud Borrego[2,3], Marcela Mena[2,3], Juan I Pieras[2,3], Mohamed F El-Ashry[1], Leen Abu Safieh[1], Amna Shah[1], Michael E Cheetham[6], Nigel P Carter[5], Christina Chakarova[1], Chris P Ponting[4], Shomi S Bhattacharya[1,7] & Guillermo Antinolo[2,3]

Using a positional cloning approach supported by comparative genomics, we have identified a previously unreported gene, *EYS*, at the RP25 locus on chromosome 6q12 commonly mutated in autosomal recessive retinitis pigmentosa. Spanning over 2 Mb, this is the largest eye-specific gene identified so far. *EYS* is independently disrupted in four other mammalian lineages, including that of rodents, but is well conserved from *Drosophila* to man and is likely to have a role in the modeling of retinal architecture.

Autosomal recessive retinitis pigmentosa is one of the most debilitating hereditary retinal disorders leading to severe visual impairment. To date 26 genetic loci have been implicated in autosomal recessive retinitis pigmentosa; however, with the exception of RP25 (MIM602772; RetNet), the prevalence of each of the reported loci is only 1–5%.

Previous genetic studies mapped the RP25 locus to a ~16-cM region on chromosome 6p12.1–q15 in four Spanish families[1]. Subsequently, linkage to the same locus was reported in multiple families from various ancestral origins, supporting RP25 as the first major locus for recessive retinitis pigmentosa[2–4].

A detailed review of the RP25 interval revealed information on 110 genes, emphasizing the extent of the work required to identify the causative gene. We therefore adopted the approach of (i) exclusion of 15 candidate genes, such as *GABRR1*, *GABRR2*, *MYO6*, *EEF1A1*, *ELOVL4*, *RIMS1*, *IMPG1* and *LCA5* (*C6ORF52*)[5], on the basis of their known retina-related function or their involvement in other retinal degenerations overlapping with RP25; and (ii) systematic screening of a further 45 genes, leading to the exclusion of 60 of the original 110 genes[6]. At this stage, mapping of five additional families at the RP25 locus helped refine the disease interval to a 2.67-cM region between D6S257 and D6S1557, thereby focusing our search for the disease-associated gene to a much smaller interval[7]. In parallel, a high-throughput screening for deletions was undertaken using array comparative genomic hybridization (array CGH). Notably, we found that a ~100-kb clone (chr6tp-19C7) was deleted in all affected members of one of the originally linked families (RP5)[6]. This suggested that the deleted clone could contain or overlap with the disease-associated gene (provisionally referred to here as the *RP25* gene). The genomic sequence spanning this deletion contained six independently predicted genes, *Q5T1H1*, *Q9H557*, *Q5TEL3*, *Q5TEL4*, *Q5VVG4* and *Q5T3C8*, which became the priority for mutation screening.

We carried out molecular analysis of these six genes in ten previously linked Spanish families, three of which are consanguineous (RP5, RP167 and RP377) and seven nonconsanguineous (RP73, RP214, RP299, RP328, RP349, RP351 and RP355)[1,7]. In one of these predicted genes, *Q9H557*, we observed a homozygous 17-bp deletion in family RP214 (Supplementary Fig. 1a online). Additionally, the four predicted exons from the same gene and a single-exon gene, *Q5TEL3*, failed to amplify in all affected members of family RP5 (Supplementary Fig. 2 online). This confirmed the ~100-kb deletion seen earlier by array CGH and strongly implicated *Q9H557* and *Q5TEL3* as part of the *RP25* gene.

To fully characterize the *RP25* gene, we carried out multiple RT-PCRs and RACE using the sequence information of *Q9H557* and *Q5TEL3* (Supplementary Methods online). This approach, combined with the comparative genomic analysis, allowed us to unravel the structure of this newly identified gene, which we eventually determined to be an unannotated prediction. On assembling all available data we noted that *RP25* encompasses 30 exons belonging to nine previously predicted genes and 13 newly reported exons, and spans the interval between 64,487,835 and 66,473,839 on chromosome 6q12 (Fig. 1). Of note, we found that a 1,238-bp segment spanning the 3′ end of exon 29 was entirely absent from the reference human genome assembly but was well represented within trace archive sequences (Fig. 1d). The full coding region of the *RP25* gene (~9 kb) was amplified in two overlapping fragments (Supplementary Fig. 3 online). RT-PCR analysis of cDNAs from a variety of normal tissues and cell lines using cDNA specific primers within *RP25* (Supplementary Table 1 online) amplified the expected-size product only from the retina and from a photoreceptor-like cell line, Y79 (Fig. 2a).

[1]Department of Molecular Genetics, Institute of Ophthalmology, London EC1V 9EL, UK. [2]Unidad de Gestión Clínica de Genética, Reproducción y Medicina Fetal, Hospitales Universitarios Virgen del Rocío, Avenida Manuel Siurot s/n, 41013 Seville, Spain. [3]CIBER de Enfermedades Raras (CIBERER), Seville, Spain. [4]Medical Research Council Functional Genomics Unit, University of Oxford, Department of Physiology, Anatomy and Genetics, Le Gros Clark Building, South Parks Road, Oxford OX1 3QX, UK. [5]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [6]Division of Molecular and Cellular Neuroscience, Institute of Ophthalmology, London EC1V 9EL, UK. [7]Institut de la Vision, INSERM U592, Université Pierre et Marie Curie-Paris 6, UMRS 592, Paris, France. [8]These authors contributed equally to this work. Correspondence should be addressed to S.S.B. (smbcssb@ucl.ac.uk) or G.A. (guillermo.antinolo.sspa@juntadeandalucia.es).

Received 20 June; accepted 19 August; published online 5 October 2008; doi:10.1038/ng.241

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

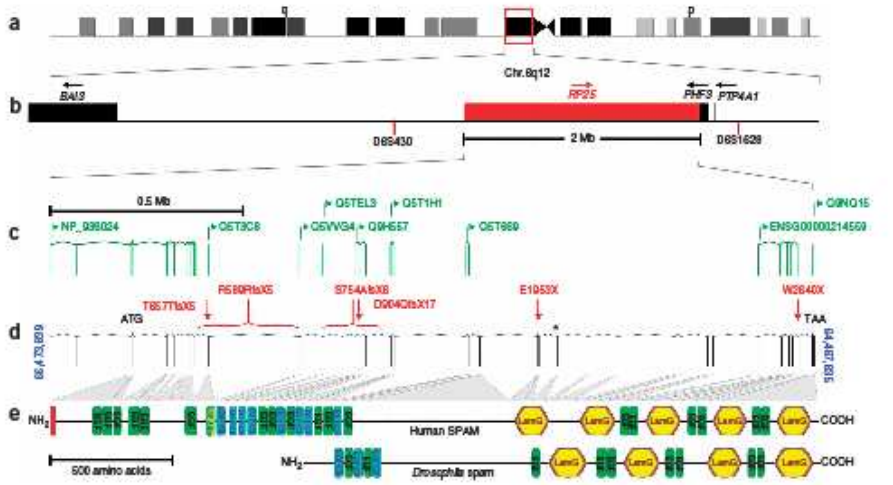


Figure 1 *RP25* gene structure and domain architecture. (a) Chromosomal region at 6q12. (b) Schematic representation of the genes and microsatellite markers flanking the *RP25* gene. (c) Previously predicted genes overlapping *RP25*. (d) The 43 exons comprising the *RP25* gene with the initiation (ATG) and stop codon (TAA) marked within exons 4 and 43, respectively; mutations are indicated in red and the asterisk at exon 29 marks the 1,238-bp segment missing from the human reference assembly. (e) Domain architecture of human SPAM and its *Drosophila* spam ortholog.

We then used a combination of methods incorporating direct sequence analysis, array CGH and the multiplex ligation-dependent probe amplification (MLPA) techniques to ensure comprehensive mutation screening of the coding regions and splice sites of the 43 exons comprising *RP25* (Supplementary Methods). So far, we have identified six independent mutations, including four deletions and two nonsense substitutions, all leading to premature stop codons in five unrelated families (Table 1 and Supplementary Note online). It is known that mRNA containing premature stop codons undergo nonsense-mediated decay[8]; therefore, the disease mechanism in these families may be due to complete absence of a functional protein.

Direct sequence analysis was not appropriate for detecting the large heterozygous deletion in family RP73 or defining the deletion breakpoints in families RP5 and RP73; hence, we used both array CGH and MLPA. We also used restriction-digest analysis as a second independent method to study the segregation of the nonsense substitutions and the small deletions (Supplementary Fig. 1a–d). None of the identified mutations have been detected in 200 control individuals.

RP25 is predicted to be a multidomain protein containing 3,145 amino acids with at least 21 epidermal growth factor (EGF)-like domains in its N-terminus followed by five C-terminal LamG domains, interspersed by further EGF repeats (Fig. 1e). This unique domain structure[9] was also seen in the *Drosophila* spacemaker (spam) protein (Supplementary Methods), encoded by *eys* (*eyes shut*). We have accordingly named the human *RP25* gene *EYS* (encoding the protein SPAM). *Drosophila* spam is expressed in the eye across diverse insect species with an open rhabdom system, such as fruitflies (*Drosophila melanogaster*) and houseflies (*Musca domestica Linnaeus*), in which the rhabdomeres or photoreceptor cells of each ommatidium

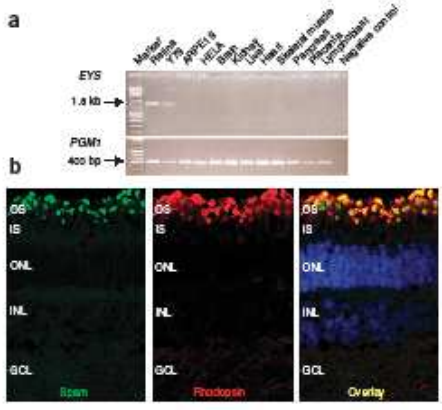


Figure 2 Expression pattern and immunolocalization of spam. (a) *EYS* (SPAM) expression in different tissues is shown in the upper panel with a specific 1.8-kb product in the retina and in Y79 photoreceptor-like cells. ARPE19 is a retinal pigment epithelial cell line. A 400-bp fragment representing the gene *PGM1*, which is ubiquitously expressed in all tissues and cell lines, is shown as an amplification control in the lower panel. (b) Subcellular localization of spam to the outer segment of mature porcine retina using antibody to spam (green) and antibody to rhodopsin (red). The overlay shows the localization of spam in the same region as rhodopsin (yellow) in the outer segment of the photoreceptor cell layer.

Table 1 Mutations identified within the *RP25* gene

Family ID	Exon	Nucleotide position	Protein alteration	Type of mutation
RP214	17	2710_2726del17	D904QfsX17	Homozygous
RP5	15–19	2260-511 91_2992+45990	S754AfsX6	Homozygous
RP328	12	1971delT	T657TfsX5	Homozygous
RP73	12	[1767-24596_2023+238135] +	R589RfsX5	Heterozygous
	28	[5857G>T]	E1953X	Heterozygous
RP349	41	7919G>A	W2640X	Homozygous

in the compound eye are separated from each other. In contrast, species with a closed system, such as the mosquito (*Anopheles gambiae*), rust-red flour beetle (*Tribolium castaneum*) and the honeybee (*Apis mellifera*), do not express spam in the eye. The complete loss of *eys* (spam) converts an open rhabdom system to a closed one, whereas its targeted expression to photoreceptors of a closed system markedly reorganizes the architecture of the compound eyes to resemble an open system[10]. On the basis of these findings in *Drosophila* and the RT-PCR data (Fig. 2a), we expected SPAM to localize in the photoreceptor layer, and indeed our immunohistochemical studies confirmed this localization (Fig. 2b).

An apparently intact *eys* gene is found across the mammalian clade, including monotremes (platypus) and marsupials (opossum) (Supplementary Fig. 4 online). However, despite the mutations and the presumed loss of function associated with human disease, this gene has been dispensed with on at least four separate occasions in the last ~100 million years of mammalian evolution[11], including in the armadillo (*Dasypus novemcinctus*), little brown bat (*Myotis lucifugus*) and ruminant (cattle and sheep) lineages (Supplementary Methods). *Eys* has acquired many (≥3) reading-frame disruptions in three rodents (mouse, rat and guinea pig) representing two of the three major rodent clades (Supplementary Fig. 4)[11,12]. This was also confirmed by failure of PCR amplification of *eys* from mouse retinal cDNA and further supported by the absence of any immunolocalization signal in the mouse retina (data not shown). *EYS* is only the fourth mendelian disease–associated human gene whose orthologs are disrupted or absent from rodent genomes[13].

In summary, we report the identification and genomic characterization of a previously unreported gene, *EYS* (encoding SPAM), implicated in autosomal recessive retinitis pigmentosa. The identification of six independent mutations and the presence of linked families from different ancestral origins support *EYS* as the first major gene reported for autosomal recessive retinitis pigmentosa. With 43 exons, covering 2.0 Mb, this is the largest gene identified to be expressed in the human eye and the fifth largest overall in the human genome. Information about the established function of insect orthologs suggests that *EYS* may possess similar functions in maintaining the integrity of the photoreceptor cells in human retina.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We would like to thank the families who participated in the study. This study was funded by Fondo de Investigación Sanitaria (PI050857), Spain; Consejería de Salud (PI-0334/2007), Junta de Andalucía, Spain; British Retinitis Pigmentosa Society (grant ref. GR556); Foresight, Dubai; Foundation Fighting Blindness (USA); National Institute of Health Research Biomedical Research Centre for Ophthalmology, The Special Trustees of Moorfields Eye Hospital London; the UK Medical Research Council and EU-Neurotrain (grant ref. MEST-CT-2005-020235); EU-GENORET (grant ref. LSHG-CT-2005-512036). The El Centro de Investigación Biomédica en Red de Enfermedades Raras is an initiative of the Instituto de Salud Carlos III. N.P.C. and E.P. were supported by the Wellcome Trust. We would like also to thank R. Molday (University of British Columbia, Vancouver) for the gift of 1D4 antibody and G. Jeffery for his advice with immunohistochemistry.

AUTHOR CONTRIBUTIONS

S.S.B. and G.A. designed the study; M.M.A., I.B., C.A.O., I.I.P., M.E.E. and L.A.S. performed the mutation screening; I.B., M.M. and S.B. performed the MLPA experiments; C.A.O., A.S., C.C. and M.E.C. performed the immunohistochemistry; L.G. and C.P.P. carried out the bioinformatics analysis and the evolutionary work; E.P. and N.P.C. carried out the array-CGH analysis; M.M.A. and S.S.B. were mainly responsible for the writing of the manuscript and prepared the tables, figures and supplementary material, with input from all coauthors.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Ruiz, A., Borrego, S., Marcos, I. & Antiñolo, G. *Am. J. Hum. Genet.* **62**, 1452–1459 (1998).
2. Khaliq, S. *et al. Am. J. Hum. Genet.* **65**, 571–574 (1999).
3. Bamgbin, I., Marcos, I., Borrego, S. & Antiñolo, G. *Ophthalmic Res.* **37**, 89–93 (2005).
4. Abd El-Aziz, M.M. *et al. Ann. Hum. Genet.* **71**, 281–299 (2007).
5. den Hollander, A.I. *et al. Nat. Genet.* **39**, 889–895 (2007).
6. Abd El-Aziz, M.M. *et al. Ann. Hum. Genet.* **72**, 463–477 (2008); published online 29 May 2008.
7. Bamgan, I. *et al. Ann. Hum. Genet.* **72**, 454–462 (2008); published online 29 May 2008.
8. Holbrook, J.A., Neu-Yilik, G., Hentze, M.W. & Kulozik, A.E. *Nat. Genet.* **36**, 801–808 (2004).
9. Husain, N. *et al. Dev. Cell* **11**, 483–493 (2006).
10. Zelhof, A.C., Hardy, R.W., Becker, A. & Zuker, C.S. *Nature* **443**, 696–699 (2006).
11. Murphy, W.J., Pringle, T.H., Crider, T.A., Springer, M.S. & Miller, W. *Genome Res.* **17**, 413–421 (2007).
12. Huchon, D. *et al. Mol. Biol. Evol.* **19**, 1053–1065 (2002).
13. Huang, H. *et al. Genome Biol.* **5**, R47 (2004).

© 2008 Nature Publishing Group http://www.nature.com/naturegenetics

human genetics

doi: 10.1111/j.1469-1809.2008.00455.x

Large-scale Molecular Analysis of a 34 Mb Interval on Chromosome 6q: Major Refinement of the RP25 Interval

M. M. Abd El-Aziz[1,5,#], I. Barragan[2,3,#], C. O'Driscoll[1], S. Borrego[2,3], L. Abu-Safieh[1], J. I. Pieras[2,3], M. F. El-Ashry[1], E. Prigmore[4], N. Carter[4], G. Antinolo[2,3] and S. S. Bhattacharya[1,*]

[1]*Department of Molecular Genetics, Institute of Ophthalmology, London EC1V 9EL, UK*
[2]*Unidad Clínica de Genética y Reproducción, Hospitales Universitarios Virgen del Rocio, Avenida Manuel Siurot s/n, 41013 Seville, Spain*
[3]*Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Seville, Spain*
[4]*The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*
[5]*Department of Clinical Pathology, Tanta University Hospital, Tanta, Egypt*

Summary

A large scale bioinformatics and molecular analysis of a 34 Mb interval on chromosome 6q12 was undertaken as part of our ongoing study to identify the gene responsible for an autosomal recessive retinitis pigmentosa (arRP) locus, RP25. Extensive bioinformatics analysis indicated in excess of 110 genes within the region and we also noted unfinished sequence on chromosome 6q in the Human Genome Database, between 58 and 61.2 Mb. Forty three genes within the RP25 interval were considered as good candidates for mutation screening. Direct sequence analysis of the selected genes in 7 Spanish families with arRP revealed a total of 244 sequence variants, of which 67 were novel but none were pathogenic. This, together with previous reports, excludes 60 genes within the interval (~55%) as disease causing for RP. To investigate if copy number variation (CNV) exists within RP25, a comparative genomic hybridization (CGH) analysis was performed on a consanguineous family. A clone from the tiling path array, chr6tp–19C7, spanning ~100–Kb was found to be deleted in all affected members of the family, leading to a major refinement of the interval. This will eventually have a significant impact on cloning of the RP25 gene.

Keywords: retinitis pigmentosa, mutation screening, gene expression

Introduction

Retinal dystrophies represent the most common inherited form of human visual handicap, with an estimated prevalence of 1 in 3000. They can be classified according to which photoreceptor system, rods or cones, is primarily affected (Bird, 1995). Disorders primarily affecting the rod system such as retinitis pigmentosa (RP) present initially with night blindness and may progress to involve the peripheral visual field, with relative preservation of central vision. On the other hand, cone dystrophies manifest initially with loss of central visual acuity and defects in colour vision without significant loss of peripheral vision.

*Corresponding author: Shomi S. Bhattacharya, Department of Molecular Genetics, Institute of Ophthalmology, 11–43 Bath Street, London EC1V 9EL, UK. Tel: +44 020 7608 6826; Fax: +44 020 7608 6863; E-mail: smbcssb@ucl.ac.uk
#Contributed equally

Several loci with retinal dystrophy phenotypes have been mapped to the pericentromeric region of chromosome 6 (6q14–q21) (Abd El-Aziz et al. 2005). The genes for most of these loci have been identified. For example, *ELOVL4* is the gene for autosomal dominant Stargardt–like disease (STGD3) as well as for autosomal dominant macular dystrophy (adMD) (Zhang et al. 2001). RIM1 and IMPG1 have also been reported as the causative genes for cone-rod dystrophy (CORD7) and benign concentric annular macular dystrophy (BCAMD), respectively (Johnson et al. 2003; van Lith-Verboeven et al. 2004). Recently, the gene for Leber congenital amaurosis type 5 (LCA5), *Lebercilin*, was also identified (den Hollander et al. 2007). It is interesting to note that the above retinal genes overlap with the autosomal recessive RP locus, RP25, for which the gene still remains to be identified.

The locus for RP25 was originally established by homozygosity mapping through targeting functional candidate genes. This resulted in the mapping of 4 Spanish

© 2008 The Authors
Journal compilation © 2008 University College London

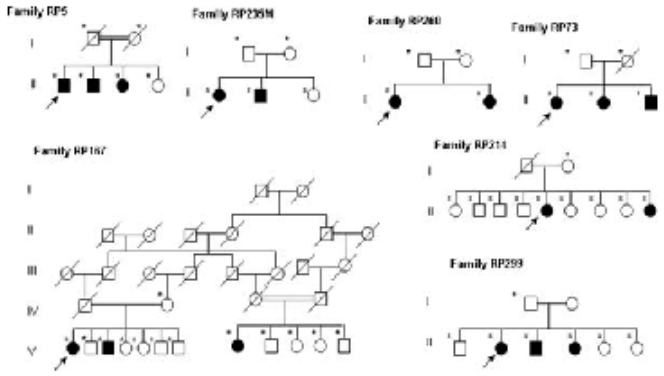


Figure 1 R.P25 families which participated in the study. Open and closed symbols denote unaffected and affected individuals, respectively. Deceased family members are denoted by diagonal slashes, asterisks indicate individuals examined both clinically and genetically and arrows indicate probands in each family.

families with arRP between microsatellite markers D6S257 and D6S1644 (Ruiz et al. 1998). The region flanked by these markers is ~34 Mb in size and contains approximately 110 known and predicted genes. Subsequently, evidence of linkage to the same region has been reported in three additional Spanish families with arRP as well as in other ethnic groups (Khaliq et al. 1999; Barragan et al. 2005a; Abd El-Aziz et al. 2007). Based on these findings, the isolation of the RP25 gene would represent a major achievement in RP research.

Previously we and others reported the exclusion of 17 genes from the RP25 interval (Marcos et al. 2000; Li et al. 2001; Marcos et al. 2002, 2003; Barragan et al. 2005a, 2005b; Abd El-Aziz et al. 2005, 2006; Barragan et al. 2008. Herein we report extensive molecular analysis of the original RP25 interval (~34 Mb) and mutation screening of a further 43 genes; therefore along with previous reports, 55% of the 110 genes have been evaluated as to their role in disease-causation for RP25. In addition, a comparative genomic hybridization (CGH) analysis was performed on one of the RP25 families to investigate whether copy number variations (CNVs) exist and if this could have an impact on the RP25 phenotype.

Materials and Methods

Families and DNA

Seven Spanish families, two consanguineous (RP5 and RP167) and five non-consanguineous (RP73, RP214, RP299, RP260 and RP235), were included in the study (Fig. 1). Informed consent was obtained from all participants for clinical and molecular genetic studies. The study conformed to the tenets of the Declaration of Helsinki.

Bioinformatics Analysis

Genomic sequence of the screened genes within the interval was accessed through the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/), the ENSEMBL database (http://www.ensembl.org) and the UCSC human genome browser (http://genome.ucsc.edu/). Information on expression pattern and expressed sequence tags (ESTs) were obtained from the NCBI UniGene database. In cases where several alternatively spliced transcripts were documented in ENSEMBL, or the exonic structure was uncharacterised, the BLAST tool (available through the NCBI) and the 6 Frame Translation tool from the Baylor College of Medicine (BCM) Search Launcher (http://searchlauncher.bcm.tmc.edu/seq-util/seq-util.html) were used to compare the sequences from these transcripts and the human genome sequence databases and to define the coding and protein sequences.

In order to evaluate the pathogenicity of the novel variants, we employed various softwares which analyse the potential role of a given variant on the function or structure of the encoded protein based on conservation and homology, physical properties of the amino acids, and prediction of the protein disorder (Conseq: http://conseq.tau.ac.il/; PolyPhen: http://coot.embl.de/PolyPhen/; SIFT: http://blocks.fhcrc.org/sift/SIFT.html; Disopred: http://bioinf.cs.ucl.ac.uk/disopred/disopred.html) (Thusberg & Vihinen 2006). In addition, intronic variants were evaluated for affecting any regulatory process at the transcriptional or splicing levels (TESS Transcription Element

© 2008 The Authors
Journal compilation © 2008 University College London

Search System: http://www.cbil.upenn.edu/cgi-bin/tess/tess; http://www.fruitfly.org/seq_tools/splice.html Splice Signal Analysis: http://www.ebi.ac.uk/asd-srv/wb.cgi; Alternative Splicing DataBase: http://hazelton.lbl.gov/~teplitski/alt/; http:/www.fruitfly.org/cgi-bin/seq_tools/splice.pl; Splicing Element Annotation: http://genes.mit.edu/acescan2/index.html; http://www.ensembl.org/Homo_sapiens/generegulationview; http://www.cisred.org/content/software; http://regrna.mbc.nctu.edu.tw/html/about.html) (Yeo et al. 2004; Matlin et al. 2005; Wang & Marin 2006).

PCR Amplification of Candidate Genes

In total 530 primers pairs were designed using Primer 3 Output (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) in order to ensure a total coverage of the entire coding region, the intronic flanking sequences, the regulatory factors binding sites and the 5′ untranslated region (UTR.) of the major transcript as well as the additional exons contributed by other published alternatively spliced isoforms (details of primers are available on request). Polymerase chain reaction (PCR.) was carried out as previously reported (Barragan et al. 2005a).

Mutation Screening

The PCR products were purified by adding Exosap [1 U shrimp alkaline phosphatase (SAP, Amersham LifeScience, Buckinghamshire, UK; ExoSAP-IT®, USB Corporation) and 1 U Exonuclease I (United States Biochemicals, Ohio, USA) to 1 μl of the PCR product and incubated at 37 °C for 15 minutes then at 80 °C for another 15 minutes to deactivate the enzyme. Sequencing reaction was then performed as previously described (Abd El-Aziz et al. 2005). Direct sequence analysis of one affected and one unaffected individual from each of the Spanish families was performed on the automated fluorescence DNA sequencer (ABI 3730, Applied Biosystems, Warrington, UK), according to manufacture's instructions. Subsequently, the data were analysed using SeqManTM 4.03 software (DNAStar, Wisconsin, USA).

Evaluation of the Identified Changes

Novel variants were tested in a matching control population by direct sequencing using the same protocol as for the mutation detection. In some other cases the changes were genotyped by fragment analysis using primers labelled at the 5′ end with 6-FAM fluorocrome. PCR products were then injected in an ABI3730 automated genotyper. Subsequently, data collection and allele identification were performed using GeneMapper software (Applied Biosystems).

Comparative Genomic Hybridization

Comparative genomic hybridization (CGH) was performed on 6 DNA samples (parents, 3 affected and 1 unaffected members) from one of the consanguineous families (RP5) using a whole genome tiling path (WGTP) array. The methods employed were as previously published (Redon et al. 2006). Two pairs of primers, Fragment 1 F and R. (ACAATCCCCTGTCATGGCTA and CCTTTCCACCTGTATATTAGGAGA) and Fragment 2 F and R. (GATCCACTGACAGCTTGCAC and TCGCAAATCTCATGTACTTCTCA), were designed at the beginning and end of a deleted clone. PCR amplification was then carried out in order to test the existence of this clone in all members of the family.

Results

Organization of the RP25 Interval: Distribution of Known/Predicted Genes

According to the genetic data obtained from the RP25 linked families, we have divided the RP25 interval into four regions (A, B, C and D; Fig. 2) which as stated earlier contains in excess of 110 genes. Region A, between microsatellite markers D6S257 and D6S1053, spans approximately 8 Mb and contains 15 genes. It is important to note that within this region a part of the human genome (Ensembl release 48), ~50 Kb between 58 and 59 Mb and ~3 Mb between 59 and 61.2 Mb on chromosome 6q, remains to be sequenced. A ~6 Mb interval between microsatellite markers D6S1053 and D6S1557 constitutes the second region (B) and contains 11 genes. The third region (C) is localised between D6S1557 and D6S421 spans only ~1 Mb and contains 7 genes. The last region, D, is the largest since it spans the rest of the RP25 locus (~18 Mb) and contains 79 genes (Fig. 2). Even though a large number of genes are included within the region, some of the sequenced intervals are devoid of genes. These regions could either represent gene desert areas in the human genome or regions where newly predicted genes might soon appear.

Strategies for Selecting Candidate Genes

A comprehensive bioinformatics analysis of the RP25 interval indicated approximately 110 genes with many of them showing retinal expression. Information about gene function, expression pattern and/or the genetic data was among the criteria for selecting candidate genes from this interval. Here we present data on 43 genes, and a description of the adopted strategies for selecting these genes is detailed below.

Based on the organization of the RP25 interval as described above, region A was evaluated because of the possibility that the Pakistani family, the 3 Chinese and 5 of the Spanish families may have a common gene. Hence,

© 2008 The Authors
Journal compilation © 2008 University College London

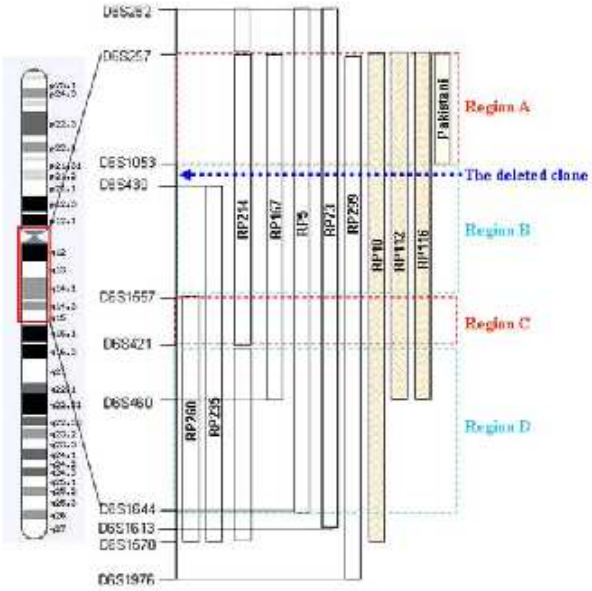


Figure 2 Schematic representation of chromosome 6, the RP25 locus (between microsatellite markers D6S257 and D6S1644). The genetic intervals for the Spanish (RP299, RP73, RP5, RP167, RP214, RP235 and RP260), Chinese (RP10, RP112 and RP116) and Pakistani families are represented by vertical bars. The dashed boxes in family RP214 represent the non homozygous linked intervals in this family. The interval is divided into four regions (A, B, C and D) according to the genetic data. The dotted arrow represents the location of the deleted clone in relation to the larger map.

12 genes from this region were considered as good candidates for mutation screening (Table 1). Secondly, 5 genes, from region B where all families except the Pakistani and one of the Spanish (RP260) families shared linkage, were also considered as high priority for screening (Fig. 2 and Table 1). Following the assumption that apart from the Pakistani family all families could have a common gene; 3 genes from region C were evaluated (Table 1).

Lastly, the possibility of the disease gene lying within region D from which the remaining 23 genes were chosen cannot be ruled out. These genes were selected according to their functional relevance to RP; for example homologous genes from the ciliated Trypanosome mapping to retinal disease loci were used to identify functional candidates (Table 1) (Broadhead et al. 2006). This was based on the hypothesis that successful transmission of several proteins from the cell body to the outer segment of the photoreceptor cells depends on their transport along a modified cilium and that defective passage of certain molecules results in retinal degeneration associated with RP (Marszalek & Goldstein, 2000). Also, the study of novel genes whose expression can be modulated by a primary defect in a retina-specific gene was also used as an approach to identify functional candidates. For this reason the genes highlighted by the retinal degeneration (rd) mouse model, caused by a defect in the β subunit of rod cGMP-phosphodiesterase, were considered as good candidates (Table 1) (Bowes et al. 1990). Finally, *EyeSAGE* (Serial Analysis of Gene Expression) and the National Eye Institute Bank (NEIBank) were used as excellent databases comparing the pattern of gene expression in the retina and retinal pigment epithelium (RPE)

© 2008 The Authors
Journal compilation © 2008 University College London

Table 1 Information about the genes screened in this study.

Gene/ Description of Gene	Gene Size (Kb) / No. of Exons / No. Primers/ Position (Mb)	UniGene	Published Information About the Gene
Genes in-between microsatellite markers D6S257 and D6S1053			
1. *BMP5* Bone morphogenetic protein 5	120.1/4/7/ 55,72–55,84	Hs.296648	Part of the transforming growth factor-beta superfamily; signaling molecule within the trabecular meshwork and optic nerve head; potential role in glaucoma pathogenesis.
2. *C6orf65/* Chromosome 6 open reading frame 65	72.05/8/10 56.92–57.00	Hs.582993	Sugar porter activity; regulation of transcription, DNA-dependent.
3. *KIAA1586/* KIAA1586 protein	8.68/4/7 57.01–57.02	Unknown	Hypothetical protein
4. *ZNF451/* Zinc finger protein 451	80.27/14/19 57.06–57.14	Hs.485628	Coactivator for steroid receptors
5. *BAG2/* BCL2-associated athanogene 2	12.61/3/4/ 57.14–57.15	HS.55220	Was reported along with BAG1 and BAG3 to interact with heat shock 70 KD protein 8 ATPase domain and inhibit its chaperone activity (Takayama et al. 1999).
6. *PRIM2A/* Primase, polypeptide 2A, 58kDa	329.44/14/11/ 57.28–57.62	HS.485640	Plays a role in both the initiation of DNA replication and the synthesis of Okazaki fragments (Shiratori et al. 1995).
7. *GUSBL2/* Glucuronidase, beta-like 2	41.64/8/9 58.35–58.39	HS.561539	Plays an important role in the degradation of dermatan and keratan sulfates.
8. *LOC442225*	2.86/1/3 63.960–63.963		Similar to KRT8 protein
9. *FKBP1C/Q5VVH2_human/* FK506 binding protein 1C	1.58/1/1 63.97–63.98	Unknown	Paralog to FKBP1A on chromosome 20p13.
10. *LOC442226*	2.86/1/3 63.99–64.00	Unknown	Similar to chromosome 15 open reading frame 2
11. *PHF3/* PHD finger protein 3	79.64/16/16 64.40–64.48	HS.348921	PHF3 expression was reported to be significantly reduced in glioma (Fischer et al. 2001).
12. *Q9NQ15/* Novel protein with EGF-like and laminin G domains – Fragment	1.2/1/3 64.488–64.489	HS.25067	Has EGF-like and laminin G domains.
Genes in-between microsatellite markers D6S1053 and D6S1557			
13. *Q5T669_human*	8.87/2/2 65.38–65.39	Unknown	Unknown
14. *EGFL11/* EGF-like-domain, multiple 11	161.09/9/11 66.10–66.26	HS.454341	Unknown
15. *BAI3/* Brain-specific angiogenesis inhibitor 3	750.22/30/31 69.40–70.15	HS.13261	Plays an important role in suppression of glioblastoma supported by its markedly reduced expression in glioblastoma cell lines (Shiratsuchi et al. 1997).

© 2008 The Authors
Journal compilation © 2008 University College London

Table 1 Continued.

Gene/ Description of Gene	Gene Size (Kb)/ No. of Exons / No. Primers/ Position (Mb)	UniGene	Published Information About the Gene
16. *LMPRD1*/C6orf209/ LMBR1 domain containing 1	21.03/16/15 70.44–70.56	Unknown	Unknown
17. *COL19A1*/ NM_001858/ Collagen, type XIX, alpha 1	345.71/51/41 70.63–70.97	HS.444842	*COL19A1* and *COL9A1*, were duplicated from the same ancestral gene of the FACIT family (Khaleduzzaman et al. 1997).
Genes in-between microsatellite markers D6S1557 and D6S421			
18. *COL9A1*/ Collagen, type IX, alpha 1	87.03/38/39/ 70.98–71.06	HS.590892	A mutation in the COL9A1 gene was reported in a family with multiple epiphyseal dysplasia (Czarny-Ratajczak et al. 2001). A homozygous mutation was also reported in a recessive form of Stickler syndrome (Van Camp et al. 2006).
19. *C6orf57*/ Chromosome 6 open reading frame 57	22.65/3/4 71.33–71.35	HS.418495	Unknown
20. *C6orf155*/ Chromosome 6 open reading frame 155	6.30/4/3 72.180–72.187	Unknown	Unknown
Genes in-between microsatellite markers D6S421 and D6S1644			
21. *KCNQ5*/ Potassium voltage-gated channel, KQT-like subfamily, member 5	573.12/14/16 73.38–73.96	HS.98129	A KCNQ5 splice variant was identified in skeletal muscle displayed altered gating kinetics (Schroeder et al. 2000).
22. *DPPA5* / Developmental pluripotency associated 5	1.23/3 74.11–74.12	Unknown	Reported to have an important role in stemness in human embryonic stem (Kim et al. 2005).
23. *C6orf150*/ Chromosome 6 open reading frame 150	27.41/6/6 74.17–74.21	HS.14577	Unknown
24. *MTO1*/ Mitochondrial translation optimisation 1 homolog	39.69/13 74.22–74.26	HS.347614	Reported to be involved in the process of mitochondrial RNA modification; an important regulatory pathway in the phenotypic expression of the deafness-associated mitochondrial A1555G mutation (Bykhovskaya et al. 2004).
25. *SLC17A5*/ Solute carrier family 17 (anion/sugar transporter), member 5	60.78/11/11 74.35–74.42	HS.597422	Defects in SLC17A5 are the cause of Salla disease (SD); an ar neurodegenerative disorders (Verheijen et al. 1999; Aula et al. 2000; Martin et al. 2003).
26. *CD109*/ CD109 molecule	128.96/33 74.46–74.59	HS.399891	A glycosylphosphatidyl -inositol (GPI)-linked cell surface antigen expressed by CD34+ acute myeloid leukemia cell lines, T-cell lines, activated T lymphoblasts and activated platelets (Lin et al. 2002). The platelet-specific antigen system, implicated in neonatal allo-immune thrombocytopenia and post-transfusion purpura, is carried by CD109 (Kelton et al. 1990).

© 2008 The Authors
Journal compilation © 2008 University College London

Table 1 Continued.

Gene/ Description of Gene	Gene Size (Kb)/ No. of Exons / No. Primers/ Position (Mb)	UniGene	Published Information About the Gene
27. *FILIP1*/ Filamin A interacting protein 1	185.69/6 76.07–76.26	HS.526972	The C terminus of the long form of rat Filip interacted with mouse filamin A (FLNA), an actin-binding protein required for cell motility (Nagano et al. 2002).
28. *SENP6* / SUMO1/sentrin specific peptidase 6	114.37/24 76.36–76.48	HS.485784	Reported to be highly expressed in reproductive organs, such as testis, ovary and prostate suggesting that it may play a role in reproduction (Kim et al. 2000).
29. *MYO6*/ Myosin VI	167.28/35/34 76.51–76.68	HS.149387	Defects in MYO6 are the cause of both ad nonsyndromic sensorineural deafness type 22 (DFNA22) (Melchionda et al. 2001) and ar congenital neurosensory deafness type 37 (DFNB37) (Ahmed et al. 2003).
30. *HTR1B*/ 5-hydroxytryptamine (serotonin) receptor 1B	1.17/1/1 78.228–78.229	HS.123016	A genetic variation (SNP; G861C) in the HTR1B gene was associated with attention deficit/hyperactivity disorder (ADHD) (Smoller et al. 2006).
31. *HMGN3*/ High mobility group nucleosomal binding domain 3	33.49/6/6 79.96–80.00	HS.77558	Unique within the HMGN family, the HMGN3 transcript undergoes alternative splicing and generates two different variants, HMGN3a and HMGN3b. (West et al. 2001).
32. *Q9H8D1_human*/ CDNA FLJ13744 fis	0.534/1/1 80.076–80.077	Unknown	Unknown
33. *C6orf152*/ Chromosome 6 open reading frame 152	52.41/8/8 80.25–80.30	HS.21945	Reported as the causative gene for Leber congenital amaurosis type 5 (LCA5) (den Hollander et al. 2007).
34. *SH3BGRL2*/ SH3 domain binding glutamic acid-rich protein like 2	72.37/4/4 80.39–80.47	HS.232772	Reported as a novel human homologue of the SH3 binding glutamic acid-rich establishing a new family of highly conserved proteins related to Thioredoxin Superfamily (Mazzocco et al. 2002).
35. *TTK*/ TTK protein kinase Dual specificity protein kinase	37.88/22/18 80.77–80.80	HS.169840	Has dual specificity protein kinase (tyrosine and serine/threonine residues) (Lindberg et al. 1993). Is required for centrosome duplication and for the normal progression of mitosis (Fisk et al. 2003).
36. *Q9H3A8*/	0.231/1/1 82.319–82.319.2	Unknown	Unknown
37. *IBTK*/ Inhibitor of Bruton agammaglobulinemia tyrosine kinase	77.49/29/29 82.93–83.00	HS.306425	Btk deficiency causes Xl agammaglobulinemia (XLA) in humans and X-linked immunodeficiency in mice (Liu et al. 2001).
38. *ME1*/ Malic enzyme 1	220.68/14 83.97–84.19	HS.21160	Computer analysis revealed the presence of additional putative recognition motifs suggesting that ME1 gene is under complex regulatory control rather than on T3 (Gonzalez-Manchon et al. 1997).

© 2008 The Authors
Journal compilation © 2008 University College London

Table 1 Continued.

Gene/ Description of Gene	Gene Size (Kb)/ No. of Exons / No. Primers/ Position (Mb)	UniGene	Published Information About the Gene
39. *CYB5R4*/ Cytochrome b5 reductase 4	100.74/16 84.62–84.72	HS.5741	Targeted inactivation of the NCB5OR gene in mice lead to severe hyperglycemia with markedly decreased serum insulin levels and nearly normal insulin tolerance at 7 weeks of age and EM showed degranulation of beta cells and hypertrophic and hyperplastic mitochondria (Xie et al. 2004).
40. *SNX14* / Sorting nexin 14	88.41/29 86.27–86.36	HS.485871	Cloned by a gene-trap strategy which set up in embryonic stem (ES) cells aiming to trap genes expressed in restricted neuronal lineages (Carroll et al. 2001).
41. *C6orf165*/ Chromosome 6 open reading frame 165	56.46/13 88.17–88.20	HS.82921	Unknown
42. *SLC35A1*/ Solute carrier family 35 (CMP-sialic acid transporter), member A1	39.36/8/8 88.23–88.27	HS.423163	By 5 prime RACE, a cDNA encoding SLC35A1, the human homolog of the murine cytidine monophosphate (CMP)-sialic acid transporter was obtained (Ishida et al. 1996).
43. *CNR1*/ Cannabinoid receptor 1	5.5/2/5 88.90–88.93	HS.75110	A triplet repeat polymorphism was significantly associated with schizophrenia especially the hebephrenic subtype (Ujike et al. 2002). A homozygous genotype within CNR1 1359A/A was associated with the vulnerability to alcohol withdrawal delirium (Schmidt et al. 2002).

Genes from 2–7, 9, 11 and 14 were considered as good candidates due to confirmed expression in eye tissues by the EyeSAGE database. Genes' numbers 41 and 42 were expressed in the cilia of the *Trypanosome* proteome. Genes' numbers 17, 33 and 43 were considered as good candidates based on the expression pattern in the retinal degeneration (rd) mouse in comparison to the wild type mouse.

to those in different tissues and hence highlighting important genes for different eye diseases including RP25 (Table 1).

Bioinformatics of the Genes within the RP25 Interval

The results of bioinformatics analyses for exon–intron structure, gene size and number of transcripts of all the screened genes apart from 3 genes, *PRIM2A* and *LOC442225* and *LOC442226*, confirmed the previously published data (Table 1). Both *LOC442225* and *LOC442226* were uncharacterised and hence the Blast and the 6 Frame Translation tools were used to identify their exonic structure. Additionally, at the time of screening the PRIM2A gene, only 10 exons were reported in the Ensembl database, the initiation codon (ATG) was predicted but the stop codon had not been identified. Recently 4 additional exons have been predicted; however the stop codon still remains to be identified and hence the open reading frame (ORF) needs to be fully characterised.

Mutation Screening

Mutation screening of the selected 43 genes led to the identification of 244 sequence variants of which 76 were novel (Table 2). All changes were assigned a nucleotide number starting at the first translation base of all genes studied according to the GenBank entries summarised in table 2.

A considerable number of novel variants were excluded as pathogenic, based on their non-segregation with the disease phenotype or their presence in the control population.

Of the 76 identified sequence variants, only 5 showed segregation with the disease phenotype and were

© 2008 The Authors
Journal compilation © 2008 University College London

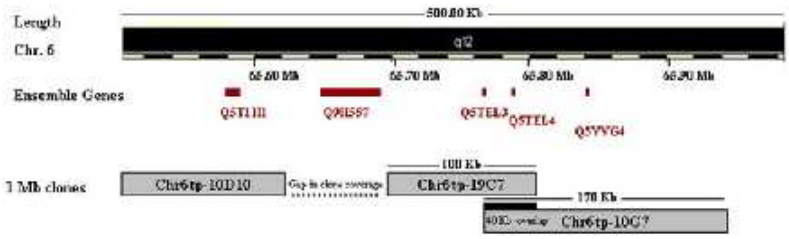


Figure 3 Schematic representation showing part of chromosome 6 along with the deleted clone, the gap in the clone coverage and the overlap between the deleted clone and the distal clone (Ch6tp-10G7). The genes within the region of the deleted clone are depicted.

Comparative Genomic Hybridization

The data obtained from the WGTP analysis revealed an inappropriate signal compatible with a deletion of a single clone in affected members of the RP5 family. The clone, chr6tp-19C7, spans ~100-Kb and is located at 65.7 Mb. A 95-Kb region proximal to this clone was not covered by the array used in this experiment. Therefore, it is possible that the deleted clone may extend over this gap. Additionally, a normal signal was observed from the Chr6tp-10D10 clone upstream of the gap. On the other hand, the deleted clone overlaps by 40-Kb with its distal neighbour, Chr6tp-10G7, which also shows a normal signal (Fig. 3). Based on this data it is possible that the deleted clone could be as much as 200-Kb in size.

The identification of at least a 100-Kb deletion in family RP5 is interesting and indicates that one of the genes, *Q5T1H1*, *Q9H557*, *Q5TEL3*, *Q5TEL4*, *Q5VVG4* and *Q5T3C8*, within this interval might be responsible for the RP25 phenotype (Fig. 3). However, we cannot rule out the possibility that this deletion might represent a rare non-pathogenic CNV. The existence of the deletion was further validated by non-amplification of two PCR amplicons designed within the deleted clone in all affected members of the family (Fig. 4). It is noteworthy that the genes within the deleted clone interval share a common epidermal growth factor (EGF-like) domain, however they represent incomplete transcripts that range between 1 to 4 exons, without an initiation or stop codon. Assuming that one of the genes from the deleted interval would be responsible for the RP25 phenotype, the families whose linked regions do not overlap with the deletion, such as RP235 and RP260 (Fig. 2), should not possess mutations in this gene. Linkage to the RP25 interval in these two families could therefore be coincidental, since many regions in the genome are expected to show suggestive evidence of linkage given the small size of these families.

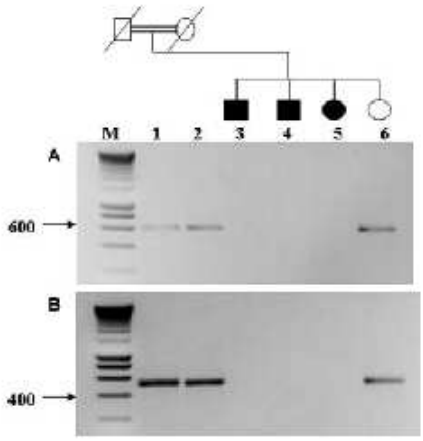


Figure 4 Cosegregation study of family RP5 using 2 pairs of primers within the deleted clone (A and B). Heterozygous parents and the unaffected daughter (lanes 1, 2 and 6) showed PCR products corresponding to the expected size of the amplified fragments. Lanes 3, 4 and 5 (affected members) showed absent PCR products confirming the data obtained from the CGH array analysis. M is a 10-Kb smart ladder.

Discussion

Herein, we report the molecular analysis of a 34 Mb interval on chromosome 6q which resulted in the exclusion of forty three genes as disease-causing for RP25 in 7 Spanish families. In addition, we have identified a ~100 Kb deletion within this interval in one of the RP25 families

© 2008 The Authors
Journal compilation © 2008 University College London

leading to a major refinement of the locus that should facilitate the subsequent identification of the disease gene (Fig. 2 and 3).

The RP25 interval spans approximately 34 Mb containing more than 110 genes, and bioinformatics analysis has revealed a sequence gap of ~3 Mb close to the centromere of chromosome 6. Moreover, some of the sequenced intervals within the RP25 region are devoid of any genes, representing either gene desert areas in the human genome or regions containing unpredicted genes that are yet to be characterised. Therefore, the number of genes within the RP25 interval should be considered as a rough estimate. However, given the large number of genes, it was impractical to screen all of them without adopting a specific strategy. As a result 43 genes were selected as good candidates for mutation screening in the Spanish families.

Both positional and functional candidate gene approaches were employed in order to identify the RP25 gene. The two strategies were overlapping throughout the study. The positional approach was considered as an essential part of the strategy for selection of candidate genes. This is due to fact that within the linkage interval there was no shared haplotype between each of the Spanish and/or other families originating from diverse geographical regions (Fig. 2). Different mutations within one gene are therefore expected to be responsible for the RP phenotype in all the families. However, the possibility of two genes responsible for the RP25 phenotype cannot be ruled out based on the presence of different crossovers from different families.

The functional candidate gene approach was also used to select genes for mutation screening. This is based on the fact that a gene could demonstrate a functional relationship with the underlying defect (RP) or there could be human homologs of genes responsible for a similar disease phenotype in other species. Alternatively, some genes might have been identified as members of a gene family of which other members have been implicated in a related disorder. For example, *C6orf165* and *SLC35A1* were considered as the human orthologs for the protein that was shown to be expressed in the cilia of the *Trypanosome* proteome, which made these genes excellent candidates (Broadhead et al. 2006).

Highly efficient genetic databases as well as information about genes expressed in the rd mouse were also used in order to identify functional candidates (Table 1). Interestingly, our strategy of selecting genes from the RP25 interval has been further validated by the identification of 2 genes within this locus, *Lebercilin* and *COL9A1*, as responsible for the retinal degeneration phenotype in both LCA5 (Leber Congenital Amaurosis), and Stickler syndrome, respectively (den Hollander et al. 2007; Van Camp et al. 2006).

Using the above described candidate gene approaches, 43 genes underwent mutation screening in 7 Spanish families. In total 244 genetic variants were identified, of which 76 were novel. Out of the 76 SNPs only 5 were initially considered as significant changes in terms of either segregation with the disease phenotype and/or absence in 192 matching control chromosomes. However, by further investigations they were ruled out as pathogenic since they have been observed in non-consanguineous families where a second change is necessary to explain their role in disease causation.

In any of the studied genes all the identified SNPs segregated with the genetic data apart from 5 genes, *PRIM2A*, *EGFL11*, *FKBP1C*, *GUSBL2* and *LOC442226*, where heterozygous SNPs were observed within an interval of homozygosity. Even though this is in contradiction to the linkage data it was only acceptable in this case since a second copy of the gene was discovered to be existing within chromosome 6q12. It has been reported that CNVs could have a direct effect on transcription regulation which in turn may influence disease susceptibility and phenotypic variation (Redon et al. 2006). Similarly, we have postulated that CNVs existing within the RP25 locus could have an impact on the phenotype of the RP25 families. For this purpose a WGTP array was carried out on one family (RP5) and the data obtained revealed that a clone within the RP25 locus was deleted in all of the affected members of the investigated family. It is interesting to note that recently 5 new Spanish families showed suggestive evidence of linkage to the RP25 locus. Moreover a crossover in one of the families has probably refined the interval from 34 to 16 Mb between microsatellite markers D6S257 and D6S1557 (published in this issue). This complements our current findings presented here.

In summary no pathogenic changes were observed in the screened genes; hence their role in the causation of the RP25 phenotype is excluded. Nevertheless, the exclusion of this significant number of genes from the interval would help in prioritising the remaining genes for mutation screening and hence in identifying the causative gene for RP25. The identification of a deleted clone from the tiling path array has narrowed down the RP25 interval from 34 Mb to only 100 Kb. It is very likely that one of the genes within the deleted clone interval could be responsible for the RP phenotype in the studied families. This confirms our initial assumption of the presence of more than one gene as responsible for RP25. If that is the case we should expect the mutations in this gene to be responsible for the RP phenotype in all RP25 linked families apart from two families, (RP235 and RP260). Future work will therefore involve full characterisation and mutation screening of the

© 2008 The Authors
Journal compilation © 2008 University College London

current and/or novel genes within the deleted interval in the RP25 families.

Acknowledgments

We would like to thank the families who participated in the study. This study was funded by Fondo de Investigación Sanitaria, Spain (PI050857), Consejería de Salud, Junta de Andalucía, Spain(PI-0334/2007), British Retinitis Pigmentosa Society (Grant ref.GR556), NIHR Biomedical Research Centre for Ophthalmology (BMRC) and Special Trustees of Moorfields Eye Hospital.

References

Abd El-Aziz, M. M., El-Ashry, M. F., Barragan, I., Marcos, I., Borrego, S., Antinolo, G. & Bhattacharya, S. S. (2005) Molecular genetic analysis of two functional candidate genes in the autosomal recessive retinitis pigmentosa, RP25, locus. *Curr Eye Res* 30, 1081–1087.

Abd El-Aziz, M. M., El-Ashry, M. F., Chan, W. M., Chong, K. L., Barragan, I., Antinolo, G., Pang, C. P. & Bhattacharya, S. S. (2007) A Novel Genetic Study of Chinese Families with Autosomal Recessive Retinitis Pigmentosa. *Ann Hum Genet* 70, 1–14.

Abd El-Aziz, M. M., Patel, R. J., El-Ashry, M. F., Barragan, I., Marcos, I., Borrego, S., Antinolo, G. & Bhattacharya, S. S. (2006) Exclusion of four candidate genes, KHDRBS2, PTP4A1, KIAA1411 and OGFRL1, as causative of autosomal recessive retinitis pigmentosa. *Ophthalmic Res* 38, 19–23.

Ahmed, Z. M., Morell, R. J., Riazuddin, S., Gropman, A., Shaukat, S., Ahmad, M. M., Mohiddin, S. A., Fananapazir, L., Caruso, R. C., Husnain, T., Khan, S. N., Riazuddin, S., Griffith, A. J., Friedman, T. B. & Wilcox, E. R. (2003) Mutations of MYO6 are associated with recessive deafness, DFNB37. *Am J Hum Genet* 72, 1315–1322.

Aula, N., Salomaki, P., Timonen, R., Verheijen, F., Mancini, G., Mansson, J. E., Aula, P. & Peltonen, L. (2000) The spectrum of SLC17A5-gene mutations resulting in free sialic acid-storage diseases indicates some genotype-phenotype correlation. *Am J Hum Genet* 67, 832–840.

Barragan, I., Borrego, S., Abd El-Aziz, M. M., El-Ashry, M. F., Abu-Safieh, L., Bhattacharya, S. S. & Antinolo, G. (2008) Genetic Analysis of FAM46A in Spanish Families with Autosomal Recessive Retinitis Pigmentosa: Characterisation of novel VNTRs. *Ann Hum Genet* 71, 281–294.

Barragan, I., Marcos, I., Borrego, S. & Antiñolo, G. (2005a) Molecular analysis of RIMI in autosomal recessive retinitis pigmentosa. *Ophthalmic Res* 37, 89–93.

Barragan, I., Marcos, I., Borrego, S. & Antinolo, G. (2005b) Mutation screening of three candidate genes, ELOVL5, SMAP1 and GLULD1 in autosomal recessive retinitis pigmentosa. *Int J Mol Med* 16, 1163–1167.

Bird, A. C. (1995) Retinal photoreceptor dystrophies LI. Edward Jackson Memorial Lecture. *Am J Ophthalmol* 119, 543–562.

Bowes, C., Li, T., Danciger, M., Baxter, L. C., Applebury, M. L. & Farber, D. B. (1990) Retinal degeneration in the rd mouse is caused by a defect in the beta subunit of rod cGMP-phosphodiesterase. *Nature* 347, 677–680.

Broadhead, R., Dawe, H. R., Farr, H., Griffiths, S., Hart, S. R., Portman, N., Shaw, M. K., Ginger, M. L., Gaskell, S. J., McKean, P. G. & Gull, K. (2006) Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature* 440, 224–227.

Bykhovskaya, Y., Mengesha, E., Wang, D., Yang, H., Estivill, X., Shohat, M. & Fischel-Ghodsian, N. (2004) Phenotype of non-syndromic deafness associated with the mitochondrial A1555G mutation is modulated by mitochondrial RNA modifying enzymes MTO1 and GTPBP3. *Mol Genet Metab* 83, 199–206.

Carroll, P., Renoncourt, Y., Gayet, O., De Bovis, B. & Alonso, S. (2001) Sorting nexin-14, a gene expressed in motoneurons trapped by an in vitro preselection method. *Dev Dyn* 221, 431–442.

Czarny-Ratajczak, M., Lohiniva, J., Rogala, P., Kozlowski, K., Perala, M., Carter, L., Spector, T. D., Kolodziej, L., Seppanen, U., Glazar, R., Krolewski, J., Latos-Bielenska, A. & Ala-Kokko, L. (2001) A mutation in COL9A1 causes multiple epiphyseal dysplasia: further evidence for locus heterogeneity. *Am J Hum Genet* 69, 969–980.

den Hollander, A. I., Koenekoop, R. K., Mohamed, M. D., Arts, H. H., Boldt, K., Towns, K. V., Sedmak, T., Beer, M., Nagel-Wolfrum, K., McKibbin, M., Dharmaraj, S., Lopez, I., Ivings, L., Williams, G. A., Springell, K., Woods, C. G., Jafri, H., Rashid, Y., Strom, T. M., van der Zwaag, B., Gosens, I., Kersten, F. F., van Wijk, E., Veltman, J. A., Zonneveld, M. N., van Beersum, S. E., Maumenee, I. H., Wolfrum, U., Cheetham, M. E., Ueffing, M., Cremers, F. P., Inglehearn, C. F. & Roepman, R. (2007) Mutations in LCA5, encoding the ciliary protein lebercilin, cause Leber congenital amaurosis. *Nat Genet* 39, 889–895.

Gonzalez-Manchon, C., Butta, N., Ferrer, M., Ayuso, M. S. & Parrilla, R. (1997) Molecular cloning and functional characterization of the human cytosolic malic enzyme promoter: thyroid hormone responsiveness. *DNA Cell Biol* 16, 533–544.

Fischer, U., Struss, A. K., Hemmer, D., Michel, A., Henn, W., Steudel, W. I. & Meese, E. (2001) PHF3 expression is frequently reduced in glioma. *Cytogenet Cell Genet* 94, 131–136.

Fisk, H. A., Mattison, C. P. & Winey, M. (2003) Human Mps1 protein kinase is required for centrosome duplication and normal mitotic progression. *Proc Nat Acad Sci* 100, 14875–14880.

Ishida, N., Miura, N., Yoshioka, S. & Kawakita, M. (1996) Molecular cloning and characterization of a novel isoform of the human UDP-galactose transporter, and of related complementary DNAs belonging to the nucleotide-sugar transporter gene family. *J Biochem* 120, 1074–1078.

Johnson, S., Halford, S., Morris, A. G., Patel, R. J., Wilkie, S. E., Hardcastle, A. J., Moore, A. T., Zhang, K. & Hunt, D. M. (2003) Genomic organisation and alternative splicing of human RIM1, a gene implicated in autosomal dominant cone-rod dystrophy (CORD7). *Genomics* 81, 304–314.

Kelton, J. G., Smith, J. W., Horsewood, P., Humbert, J. R., Hayward, C. P. M. & Warkentin, T. E. (1990) Gov(a/b) alloantigen system on human platelets. *Blood* 75, 2172–2176.

Khaliq, S., Hameed, A., Ismail, M., Mehdi, S. Q., Bessant, D. A., Payne, A. M. & Bhattacharya, S. S. (1999) Refinement of the locus for autosomal recessive Retinitis pigmentosa (RP25) linked to chromosome 6q in a family of Pakistani origin. *Am J Hum Genet* 65, 571–574.

Khaleduzzaman, M., Sumiyoshi, H., Ueki, Y., Inoguchi, K., Ninomiya, Y. & Yoshioka, H. (1997) Structure of the human type XIX collagen (COL19A1) gene, which suggests it has arisen from an ancestor gene of the FACIT family. *Genomics* 45, 304–312.

Kim, K. I., Baek, S. H., Jeon, Y. J., Nishimori, S., Suzuki, T., Uchida, S., Shimbara, N., Saitoh, H., Tanaka, K. & Chung, C. H. (2000) A new SUMO-1-specific protease, SUSP1, that is highly expressed in reproductive organs. *J Biol Chem* 275, 14102–14106.

© 2008 The Authors
Journal compilation © 2008 University College London

Kim, S. K., Suh, M. R., Yoon, H. S., Lee, J. B., Oh, S. K., Moon, S. Y., Moon, S. H., Lee, J. Y., Hwang, J. H., Cho, W. J. & Kim, K. S. (2005) Identification of developmental pluripotency associated 5 expression in human pluripotent stem cells. *Stem Cells* 23, 458–462.

Li, Y., Marcos, I., Borrego, S., Yu, Z., Zhang, K. & Antinolo, G. (2001) Evaluation of the ELOVL4 gene in families with retinitis pigmentosa linked to the RP25 locus. *J Med Genet* 38, 478–480.

Lin, M., Sutherland, D. R., Horsfall, W., Totty, N., Yeo, E., Nayar, R., Wu, X. F. and Schuh, A. C. (2002) Cell surface antigen CD109 is a novel member of the alpha-2 macroglobulin/C3, C4, C5 family of thioester-containing proteins. *Blood* 99, 1683–1691.

Lindberg, R. A., Fischer, W. H. & Hunter, T. (1993) Characterization of a human protein threonine kinase isolated by screening an expression library with antibodies to phosphotyrosine. *Oncogene* 8, 351–359.

Liu, W., Quinto, I., Chen, X., Palmieri, C., Rabin, R. L., Schwartz, O. M., Nelson, D. L. & Scala, G. (2001) Direct inhibition of Bruton's tyrosine kinase by IBtk, a Btk-binding protein. *Nature Immun* 2, 939–946.

Marcos, I., Borrego, S. & Antinolo, G. (2003) Molecular cloning and characterization of human RAB23, a member of the group of Rab GTPases. *Int J Mol Med* 12, 983–987.

Marcos, I., Galan, J. J., Borrego, S. & Antinolo, G. (2002) Cloning, characterization, and chromosome mapping of the human GlcAT-S gene. *J Hum Genet* 47, 677–680.

Marcos, I., Ruiz, A., Blaschak, C. J., Borrego, S., Cutting, G. R. & Antiñolo, G. (2000) Mutation analysis of GABRR1 and GABRR2 in autosomal recessive retinitis pigmentosa (RP25). *J Med Genet* 37, E5.

Marszalek, J. R. & Goldstein, L. S. (2000) Understanding the functions of kinesin-II. *Biochim Biophys Acta* 1496, 142–150.

Martin, R. A., Slaugh, R., Natowicz, M., Pearlman, K., Orvisky, E., Krasnewich, D., Kleta, R., Huizing, M. & Gahl, W. A. (2003) Sialic acid storage disease of the Salla phenotype in American monozygous twin female sibs. *Am J Med Genet* 120, 23–27.

Matlin, A. J., Clark, F. & Smith, C. W. (2005) Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* 6, 386–398.

Mazzocco, M., Maffei, M., Egeo, A., Vergano, A., Arrigo, P., Di Lisi, R., Ghiotto, F. & Scartezzini, P. (2002) The identification of a novel human homologue of the SH3 binding glutamic acid-rich (SH3BGR) gene establishes a new family of highly conserved small proteins related to Thioredoxin Superfamily. *Gene* 29, 233–239.

Melchionda, S., Ahituv, N., Bisceglia, L., Sobe, T., Glaser, F., Rabionet, R., Arbones, M. L., Notarangelo, A., Di Iorio, E., Carella, M., Zelante, L., Estivill, X., Avraham, K. B. & Gasparini, P. (2001) MYO6, the human homologue of the gene responsible for deafness in Snell's Waltzer mice, is mutated in autosomal dominant nonsyndromic hearing loss. *A J Hum Genet* 69, 635–640.

Nagano, T., Yoneda, T., Hatanaka, Y., Kubota, C., Murakami, F. & Sato, M. (2002) Filamin A-interacting protein (FILIP) regulates cortical cell migration out of the ventricular zone. *Nature Cell Biol* 4, 495–501.

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H., Shapero, M. H., Carson, A. R., Chen, W., Cho, E. K., Dallaire, S., Freeman, J. L., Gonzalez, J. R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J. R., Marshall, C. R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M. J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D. F., Estivill, X., Tyler-Smith, C., Carter, N. P., Aburatani, H., Lee, C., Jones, K. W., Scherer, S. W. & Hurles, M. E. (2006) Global variation in copy number in the human genome. *Nature* 444, 444–454.

Ruiz, A., Borrego, S., Marcos, I. & Antinolo, G. (1998) A major locus for autosomal recessive retinitis pigmentosa on 6q, determined by homozygosity mapping of chromosomal regions that contain gamma-aminobutyric acid-receptor clusters. *Am J Hum Genet* 62, 1452–1459.

Schmidt, L. G., Samochowiec, J., Finckh, U., Fiszer-Piosik, E., Horodnicki, J., Wendel, B., Rommelspacher, H. & Hoehe, M. R. (2002) Association of a CB1 cannabinoid receptor gene (CNR1) polymorphism with severe alcohol dependence. *Drug Alcohol Depend* 65, 221–224.

Schroeder, B. C., Hechenberger, M., Weinreich, F., Kubisch, C. & Jentsch, T. J. (2000) KCNQ5, a novel potassium channel broadly expressed in brain, mediates M-type currents. *J Biol Chem* 275, 24089–24095.

Shiratori, A., Okumura, K., Nogami, M., Taguchi, H., Onozaki, T., Inoue, T., Ando, T., Shibata, T., Izumi, M., Miyazawa, H., Hanaoka, F., Murakami, Y. & Eki, T. (1995) Assignment of the 49-kDa (PRIM1) and 58-kDa (PRIM2A and PRIM2B) subunit genes of the human DNA primase to chromosome bands 1q44 and 6p11.1-p12. *Genomics* 28, 350–353.

Shiratsuchi, T., Nishimori, H., Ichise, H., Nakamura, Y. & Tokino, T. (1997) Cloning and characterization of BAI2 and BAI3, novel genes homologous to brain-specific angiogenesis inhibitor 1 (BAI1). *Cytogenet Cell Genet* 79, 103–108.

Smoller, J. W., Biederman, J., Arbeitman, L., Doyle, A. E., Fagerness, J., Perlis, R. H., Sklar, P. & Faraone, S. V. (2006) Association between the 5HT1B receptor gene (HTR1B) and the inattentive subtype of ADHD. *Biol Psychiat* 59, 460–467.

Takayama, S., Xie, Z. & Reed, J. C. (1999) An evolutionarily conserved family of Hsp70/Hsc70 molecular chaperone regulators. *J Biol Chem* 274, 781–786.

Thusberg, J. & Vihinen, M. (2006) Bioinformatic analysis of protein structure-function relationships: case study of leukocyte elastase (ELA2) missense mutations. *Hum Mutat* 27, 1230–1243.

Ujike, H., Takaki, M., Nakata, K., Tanaka, Y., Takeda, T., Kodama, M., Fujiwara, Y., Sakai, A. & Kuroda, S. (2002) CNR1, central cannabinoid receptor gene, associated with susceptibility to hebephrenic schizophrenia. *Mol. Psychiatry* 7, 515–518.

Van Camp, G., Snoeckx, R. L., Hilgert, N., van den Ende, J., Fukuoka, H., Wagatsuma, M., Suzuki, H., Smets, R. M. E., Vanhoenacker, F., Declau, F., Van De Heyning, P. & Usami, S. (2006) A new autosomal recessive form of Stickler syndrome is caused by a mutation in the COL9A1 gene. *Am J Hum Genet* 79, 449–457.

van Lith-Verboeven, J. J. C., Hoyng, C. B., van den Helm, B., Deutman, A. F., Brink, H. M. A., Kemperman, M. H., de Jong, W. H. M., Kremer, H. and Cremers, F. P. M. (2004) The benign concentric annular macular dystrophy locus maps to 6p12.3-q16. *Invest Ophthalmol Vis Sci* 45, 30–35.

Verheijen, F. W., Verbeek, E., Aula, N., Beerens, C. E. M. T., Havelaar, A. C., Joosse, M., Peltonen, L., Aula, P., Galjaard, H., Van Der Spek, P. J. & Mancini, G. M. S. (1999) A new gene, encoding an anion transporter, is mutated in sialic acid storage diseases. *Nat Genet* 23, 462–465.

Wang, M. & Marin, A. (2006) Characterization and prediction of alternative splice sites. *Gene* 366, 219–227.

West, K. L., Ito, Y., Birger, Y., Postnikov, Y., Shirakawa, H. & Bustin, M. (2001) HMGN3a and HMGN3b, two protein isoforms with a tissue-specific expression pattern, expand the cellular repertoire of nucleosome-binding proteins. *J Biol Chem* 276, 25959–25969.

Xie, J., Zhu, H., Larade, K., Ladoux, A., Seguritan, A., Chu, M., Ito, S., Bronson, R. T., Leiter, E. H., Zhang, C. Y., Rosen, E. D.

© 2008 The Authors
Journal compilation © 2008 University College London

& Bunn, H. F. (2004) Absence of a reductase, NCB5OR, causes insulin-deficient diabetes. *Proc Nat Acad Sci* **101**, 10750–10755.
Yeo, G., Hoon, S., Venkatesh, B. & Burge, C. B. (2004) Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc Natl Acad Sci U S A* **101**, 15700–15705.
Zhang, K., Kniazeva, M., Han, M., Li, W., Yu, Z., Yang, Z., Li, Y., Metzker, M. L., Allikmets, R., Zack, D. J., Kakuk, L. E., Lagali, P. S., Wong, P. W., MacDonald, I. M., Sieving, P. A., Figueroa, D. J., Austin, C. P., Gould, R. J., Ayyagari, R. & Petrukhin, K. (2001) A 5-bp deletion in ELOVL4 is associated with two related forms of autosomal dominant macular dystrophy. *Nat Genet* **27**, 89–93.

Received: 25 January 2008
Accepted: 21 April 2008

© 2008 The Authors
Journal compilation © 2008 University College London

annals of human genetics

doi: 10.1111/j.1469-1809.2008.00448.x

Linkage Validation of *RP25* Using the 10K GeneChip Array and Further Refinement of the Locus by New Linked Families

I. Barragán[1,2,#], M. M. Abd El-Aziz[3,4,#], S. Borrego[1,2], M. F. El-Ashry[3,5], C. O'Driscoll[3], S. S. Bhattacharya[3] and G. Antiñolo[1,2,*]

[1]*Unidad Clínica de Genética y Reproducción, Hospitales Universitarios Virgen del Rocío, Seville, Spain*
[2]*Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Seville, Spain*
[3]*Department of Molecular Genetics, Institute of Ophthalmology, London EC1V 9EL, UK*
[4]*Department of Clinical Pathology, Tanta University Hospital, Tanta, Egypt*
[5]*Department of Ophthalmology, Tanta University Hospital, Tanta, Egypt*

Summary

Retinitis pigmentosa (RP) is a clinically and genetically heterogeneous group of retinal dystrophies, characterised by rod photoreceptor cell degeneration with autosomal recessive RP (arRP) as the commonest form worldwide. To date, a total of 26 loci have been reported for arRP, each having a prevalence of 1–5%, except for the *RP25* locus which was identified as the genetic cause of 14% of arRP cases in Spain. In order to validate the original linkage of *RP25*, we undertook a total genome scan using the 10K GeneChip mapping array on three of the previously linked families. The data obtained supported the initial findings of linkage. Additionally, linkage analysis in 18 newly ascertained arRP families was performed using microsatellite markers spanning the chromosome 6p12.1–q15 interval. Five out of the 18 families showed suggestive evidence of linkage to *RP25*, hence supporting the high prevalence of this locus in the Spanish population. Furthermore, the finding of a crossover in one of these families is likely to have refined the disease interval from the original 16 cM to only a 2.67 cM region between *D6S257* and *D6S1557*.

Keywords: Retinitis pigmentosa, *RP25* locus, retinal disease gene, chromosome 6

Introduction

Retinitis pigmentosa (RP) is a clinically and genetically heterogeneous group of retinal dystrophies, affecting approximately 1 in 4000 individuals worldwide (Ingleheam 1998). Patients typically present with night blindness and constriction of peripheral visual fields due to degeneration of rod photoreceptor cells. Clinical manifestations include waxy pallor of the optic disc, attenuation of retinal blood vessels with later atrophy of the retinal pigment epithelium. Electroretinographic (ERG) changes are present with abnormalities of both rod and cone ERGs, with the scotopic alteration the first to be reported. However, in advanced RP both rod and cone ERG responses are undetectable (Jiménez-Sierra et al. 1989).

Notable allelic and non-allelic heterogeneity are characteristic of RP and the disease can present with different modes of inheritance, such as X-linked (xl), autosomal dominant (ad), autosomal recessive (ar) and digenic. The autosomal recessive form of RP is the commonest worldwide, accounting for approximately 39% of cases in Spain (Ayuso et al. 1995). To date, 26 loci have been reported as responsible for arRP, of which 21 genes have been identified (http://www.sph.uth.tm.edu/Retnet/). However, all together the reported loci are responsible for only ~10–15% of the recessive RP cases (Tuson et al. 2004). This is in comparison to the *RP25* locus which was identified as the genetic cause of 10–20% of arRP cases in Spain (Ruiz et al. 1998; Woods et al. 2006).

*Corresponding author: Guillermo Antiñolo, MD, PhD, Unidad Clínica de Genética y Reproducción, Hospitales Universitarios Virgen del Rocío, Avenida Manuel Siurot s/n, 41013 Seville, Spain. Tel: +34 95 50127 72; Fax: +34 95 501 34 73; E-mail: guillermo.antinolo.sspa@juntadeandalucia.es
#Contributed equally

© 2008 The Authors
Journal compilation © 2008 University College London

RP25 was originally identified by homozygosity mapping through targeting functional candidate genes. This resulted in mapping of 4 Spanish families with arRP to a 16.1 cM interval between microsatellite markers *D6S257* and *D6S1644* (Ruiz et al. 1998). Subsequently, evidence of linkage to the same region has been reported in three additional Spanish families with arRP, as well as in other ethnic groups (Khaliq et al. 1999; Barragan et al. 2005a; Abd El-Aziz et al. 2007). The isolation of the *RP25* gene would therefore represent a major achievement in RP research. Interestingly, several loci responsible for various types of retinal dystrophies have been reported to co-localise with *RP25*, among which is included Leber congenital amaurosis (LCA5), a severe form of retinal degeneration for which the gene has recently been identified (Dharmaraj et al. 2000; den Hollander et al. 2007).

Extensive bioinformatics analysis and mutation screening of candidate genes within the *RP25* interval led to the exclusion of 17 genes, *GABRR1, GABRR2, ELOVL4, GlcAT-S, TRAP2B, RAB23, RIM1, ELOVL5, GLULD1, SMAP1, EEF1A1, IMPG1, KHDRBS2, PTP4A1, KIAA1411, OGFRL1* and *FAM46A*, as disease causing (Marcos et al. 2000, 2002, 2003; Li et al. 2001; Barragan et al. 2005a; Barragan et al. 2005b; Abd El-Aziz et al. 2005, 2006; Barragan et al. 2007). This together with our recent work involving large scale bioinformatics analysis of the *RP25* interval and mutation screening of additional 43 genes ruled out ~55% of the genes from the interval as responsible for RP (submitted for publication).

The considerable technological advances in genotyping and statistical analysis of genetic markers within pedigrees are the basis of the development of genomewide linkage studies for the elucidation of causative genes (Zhang et al. 2005; Zhao et al. 2005). The initial genomewide linkage analyses employed panels of about 400 microsatellites at 10-cM average intermarker distance (Stambolian et al. 2004). However, recent progress in Single Nucleotide Polymorphism (SNP) discovery and genotyping has provided the opportunity to use this type of marker for linkage analyses as well (Sellick et al. 2004). Besides, various studies highlighted the advantageous use of SNPs compared to microsatellite markers for linkage purposes (Evans & Cardon 2004; Middleton et al. 2004; John et al. 2004).

Herein we report the validation of linkage to the *RP25* locus by genomewide linkage analysis using the 10K SNP GeneChip array in 3 previously mapped families. Additionally, linkage analysis in 18 newly ascertained arRP families was performed using microsatellite markers spanning the chromosome 6p12.1-q15 interval.

Materials and Methods

Subjects and DNA

This study involved 21 Spanish arRP families comprising 34 affected and 68 unaffected individuals, all derived from Ophthalmology Services to our Genetic Service (UCGR) (Fig. 1). The studied families included three of the originally *RP25* linked families, RP5, RP214 and RP299, and the remaining eighteen were newly ascertained.

A group of matching control individuals was also recruited. Informed consent was obtained from all participants and the study conformed to the tenets of the Declaration of Helsinki (June 1964). All subjects underwent a peripheral blood extraction for genomic DNA extraction from leukocytes using the standard protocols (Dracopoli et al. 1994).

Genotyping Using GeneChip® Human Mapping 10K Array

The three previously studied families, RP5, RP214 and RP299, were subjected to a total genome scan using the GeneChip Human Mapping 10K array in order to validate the original linkage data (Affymetrix, High Wycombe, UK). DNA samples from six affected (RP5 II-1, II-2, II-3; RP214 II-5; RP299 II-2, II-3) and 5 unaffected (RP5 I-1, I-2, II-4; RP214 I-2; RP299 I-1) members from the 3 families (Fig. 1) were genotyped using the 10K Array Xba 142 2.0, (Affymetrix, High Wycombe, UK), employing the methodology described elsewhere (Sellick et al. 2004; Abd El-Aziz et al. 2007).

The selection of family members to be genotyped with the 10K array was based on the previous linkage data (Ruiz et al. 1998; Barragan et al. 2005a). For example, in the case of family RP214, two individuals (the available parent I-2 and affected offspring II-5) were sufficient to validate the original linkage of *RP25* based on homozygosity mapping. Similarly, for family RP299 (non-consanguineous), by using one parent (I-1) and two affected offspring (II-2 and II-3) linkage to *RP25* was established based on identifying the largest region of shared haplotype between the affected members.

Analysis of the 10K Array Data

The SNP data were analysed using ExcludeAR.3, AR.1liteCHIP2 macros and the ALOHOMORA program for RP5, RP214 and RP299 families, respectively (Woods et al. 2004; Ruschendorf & Nurnberg, 2005).

Autozygosity mapping was suggested as the method of choice for identifying recessive loci using the 10K array (Lander and Botstein, 1987). Hence, ExcludeAR.3 and AR.1 liteCHIP2 macros were utilised to analyse the three affected members and the affected member in both RP5 (consanguineous) and RP214 families (non-consanguineous), respectively. Even though

© 2008 The Authors
Journal compilation © 2008 University College London

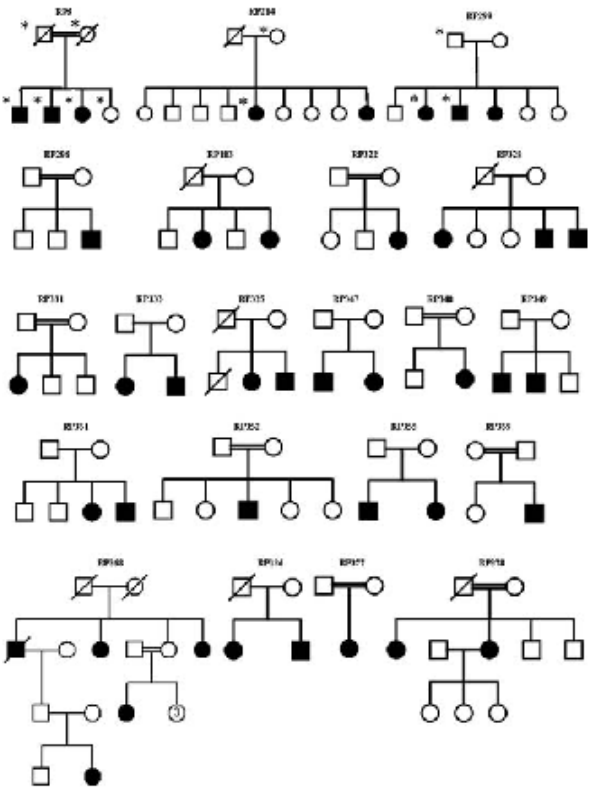


Figure 1 Shows family pedigrees of individuals who were included in the study. Open and closed symbols denote unaffected and affected individuals respectively. Deceased family members are denoted by diagonal slashes and asterisks indicate individuals that were analysed using the GeneChip 10K array.

family R.P214 was not described as consanguineous, the original linkage data revealed a region of homozygosity compatible with linkage that was shared between its 2 affected members. Therefore we have used the ExcludeAR.1 program to analyse the 10K array data from this family. Upon identifying the disease gene for R.P25, we expect the mutation in this family to be homozygous.

For family R.P299 (non-consanguineous) the ExcludeAR software was not suitable for analysing the linkage data since it identifies only homozygous intervals. Therefore, the ALOHOMORA program was used to identify genetic intervals of shared haplotype between the two affected members of this family. We have also utilised the ALOHOMORA software to analyse family R.P5 in order to confirm the data obtained from the ExcludeAR.3 macro. However, family R.P214 could not be analysed by using this method since only one affected member was available and consequently the pedigree was uninformative for comparison of haplotypes among affected members as well as for LOD score calculations.

Within ALOHOMORA, Mendelian errors and the correct relationships within families were checked for by using the PedCheck and GRR programs, respectively (O'Connell and Weeks, 1998; Abecasis et al. 2001a). Identification and deletion of non-Mendelian errors and unlikely genotypes were performed. Uninformative SNPs were removed from the data by Merlin (Abecasis

© 2008 The Authors
Journal compilation © 2008 University College London

et al. 2001b). For parametric linkage analysis, the data was then converted for Genehunter where the analysis was done in subsets of markers in a non-overlapping moving window. The haplotyping data were then used from the Genehunter file.

The analysis of the 10K array data using any of the computer programs mentioned earlier is based on identifying regions of shared haplotype between the affected members in the case of the R.P5 and R.P299 families or observing significant regions of homozygosity as for the R.P214 family. The breakpoints were identified based on the observation of a crossover between the affected members in the R.P5 and R.P299 families or the loss of homozygosity in the case of the R.P214 family. This was then followed by combining the data from the three families in order to identify significant regions of shared haplotype.

Genotyping and Microsatellite Marker Analysis of the *RP25* Locus

A total of 93 members from the newly ascertained arR.P families (Fig. 1) were subjected to genotyping analysis using microsatellite markers covering the *RP25* region. For this purpose 19 microsatellite markers were utilised, of which 11 were selected from the ABI PRISM® Linkage Mapping Sets V2.5 (Applied Biosystems, Madrid, Spain) and the remaining 8 were synthesised commercially according to the information obtained from the Ensembl database (www.ensembl.org).

Multiplex PCR amplification was employed for genotyping the microsatellite markers using genomic DNA templates and the Qiagen Multiplex PCR Kit (Qiagen, Barcelona, Spain). PCR products were injected in an automated ABI-3730 sequencer, data collection and allele identification were performed using GeneMapper software (Applied Biosystems). Multipoint parametric linkage analysis was then carried out with GENEHUNTER Complete Linkage Analysis version 2.1_r3beta software (http://linkage.rockefeller.edu/soft/gh/). The phenotype was analysed as an autosomal recessive trait with full penetrance and with a disease-gene allele frequency of 0.0001.

Results

Genomewide Linkage Analysis of the Originally Linked *RP25* Families

The data obtained from ExcludeAR3 and AR1liteCHIP2 macros for families R.P5 and R.P214 revealed the largest region of homozygosity spanning the *RP25* interval across the human genome (Table 1). This is supportive of the

Table 1 Regions of homozygosity obtained using macros in R.P5 family (ExcludeAR3) and R.P214 family (ExcludeAR1). Regions are marked as 'significant' by the program if the random chance of the observed number of consecutive SNPs in the stated genetic distance being homozygous (for AR1) combined with the chance of them being concordant in the affected siblings (for AR3) is less than one in 1000 as described in Woods et al. 2004.

R.P5 Homozygous Region by Size	cM	Chrom	SNPs	Significant?	Start cM	Finish cM
largest	50.3	6	188	**significant**	73	124
second	17.3	11	6	not significant	91	108
third	16.6	9	7	not significant	169	185
fourth	15.0	5	14	significant	187	202
fifth	9.9	20	9	not significant	3	13
sixth	9.7	10	9	not significant	112	122
seventh	7.9	5	4	not significant	239	247
eighth	7.8	2	12	significant	101	109
ninth	7.4	4	5	not significant	65	72
tenth	7.3	19	6	not significant	102	110
R.P214 **Homozygous Region by Size**	**cM**	**Chrom**	**SNPs**	**Significant?**	**Start cM**	**Finish cM**
largest	28.2	6	107	**significant**	80	112
second	24.2	2	30	not significant	301	326
third	17.8	1	44	not significant	193	211
fourth	14.6	10	13	not significant	107	121
fifth	12.9	6	11	not significant	70	83
sixth	12.7	16	11	not significant	134	146
seventh	12.0	17	10	not significant	73	85
eighth	11.1	9	4	not significant	174	185
ninth	10.8	7	9	not significant	108	119
tenth	10.5	22	7	not significant	22	32

© 2008 The Authors
Journal compilation © 2008 University College London

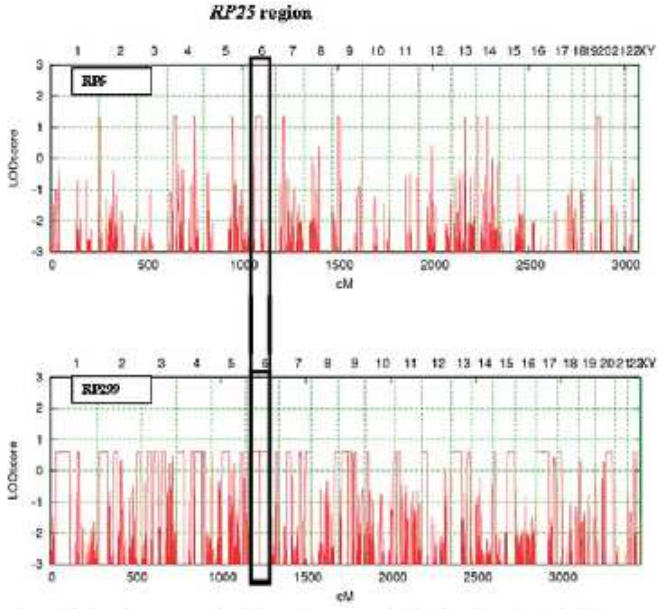


Figure 2 Schematic representation of the total genome analysis for families RP5 and RP299 generated by the ALOHOMORA software. The genetic interval (cM)/each chromosome (1–23) on the *x* axis is plotted against the LOD score on the *y* axis. The red peaks on each chromosome represent the LOD score for each of the single nucleotide polymorphisms (SNPs) used. The higher and the greater the width of the red peak the more likely it is to represent the true locus for linkage. The area enclosed by bold lines shows the RP25 locus as the largest region across the whole genome common to both families in which haplotypes within each family are identical in affected members.

original linkage data in these families (Ruiz et al. 1998). Table 1 is generated by ExcludeAR software which lists the homozygous intervals across the human genome in a descending order based on the number of SNPs and the genetic size of each interval. The *RP25* locus was considered as the disease interval in both RP5 and RP214 families since 188 and 107 SNPs were observed to be homozygous in both families, respectively.

Additionally, the data obtained from the ALOHOMORA software for both RP299 and RP5 families, highlighted the *RP25* locus as the region of shared haplotype between the affected members of both families across the human genome (Fig. 2).

Therefore, the GeneChip mapping data generated from genotyping of the three previously linked families confirmed the initial findings of linkage to the *RP25* region.

However, this did not lead to further refinement of the *RP25* genetic interval and a region of 16 cM interval has been confirmed as the disease interval in RP25 families.

Targeted *RP25* Linkage Analysis in arRP Families

Genetic linkage to *RP25* was identified in five, RP349, RP351, RP355, RP377 and RP328, out of the 18 newly ascertained arRP families (Fig. 3). In these five families, all the studied microsatellite markers in the vicinity of *RP25* generated positive LOD scores. We pooled the linkage data from all families except for RP377 and a maximum LOD score of 3.63 for microsatellite marker D6S1573 at $\theta = 0$ was generated (Table 2, Fig. 4). The reason behind

© 2008 The Authors
Journal compilation © 2008 University College London

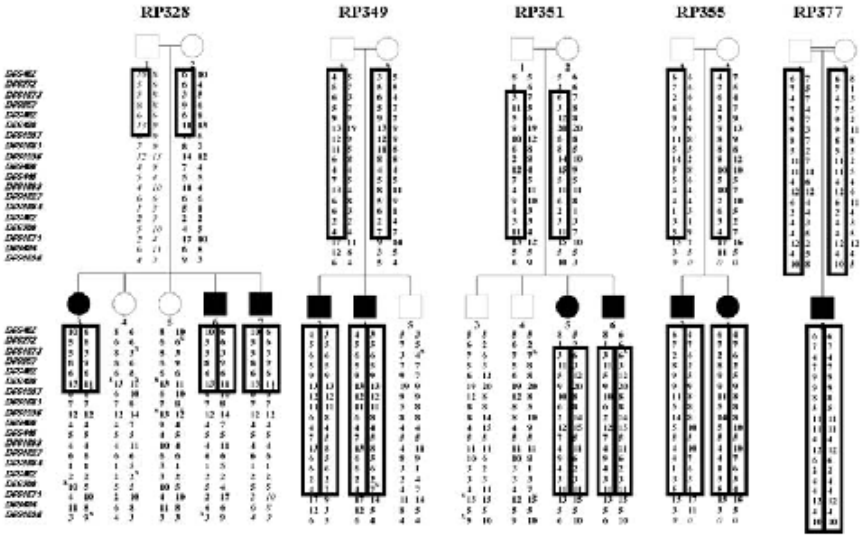


Figure 3 Haplotypes of families RP328, RP349, RP351, RP355 and RP377 using microsatellite markers on 6p12.1-q12. Recombination events between markers *D6S272* and *D6S1557* are shown in families RP351 and RP328, respectively.

Table 2 LOD score values of the individual and combined linkage data for families RP328, RP349, RP351 and RP355. Combined LOD scores have been calculated at a theta value of 0.

		LOD Score/FAMILY			
Marker	Combined LOD Score/p-value	RP328	RP349	RP351	RP355
D6S452	−∞ /0.013	1.25	0.73	−∞	0.56
D6S272	−∞ /0.007	1.37	0.71	−∞	0.60
D6S1573	**3.63/0.001**	1.45	0.73	0.85	0.60
D6S257	**3.47/ 0.001**	1.29	0.73	0.85	0.60
D6S402	**3.17/ 0.001**	1.00	0.72	0.85	0.60
D6S430	**2.82/ 0.001**	0.65	0.73	0.85	0.59
D6S1557	−∞ /0.013	−∞	0.73	0.85	0.60
D6S1681	−∞ /0.013	−∞	0.73	0.85	0.60
D6S1596	−∞ /0.013	−∞	0.73	0.85	0.60
D6S460	−∞ /0.013	−∞	0.72	0.85	0.59
D6S445	0.47/ 0.013	−1.7	0.73	0.85	0.59
D6S1609	−∞ /0.013	−∞	0.73	0.85	0.60
D6S1627	0.08/ 0.013	−2.1	0.73	0.85	0.60
D6S1004	−∞ /0.013	−∞	0.73	0.85	0.60
D6S462	0.47/0.013	−1.7	0.72	0.85	0.60
D6S300	−∞ /0.071	−∞	0.50	0.85	0.60
D6S1671	−∞ /0.495	−1.9	−∞	−∞	−∞
D6S434	−∞ /0.495	−1.6	−∞	0.09	−∞
D6S1698	−∞ /0.246	−0.87	−∞	0.85	−0.29

© 2008 The Authors
Journal compilation © 2008 University College London

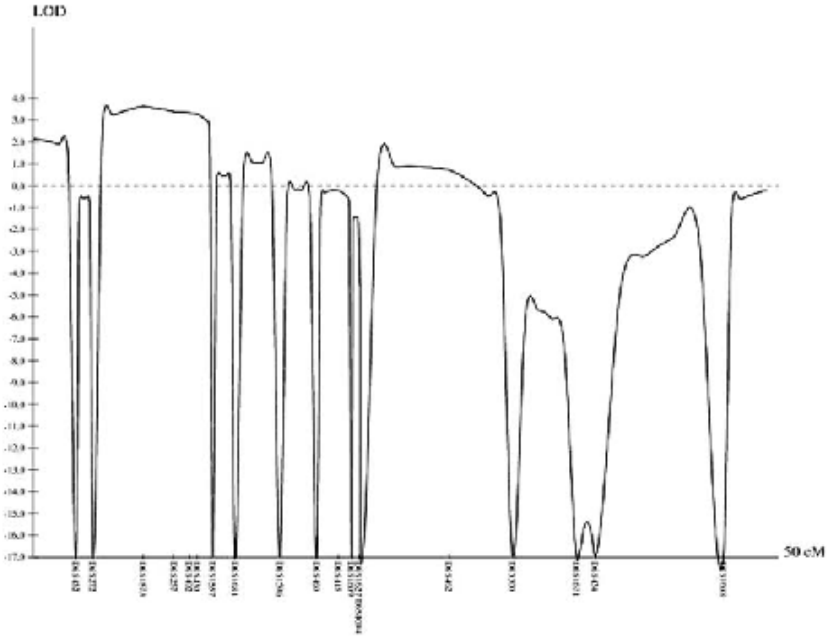


Figure 4 Graphic depiction of the LOD score variation with genetic distance calculated using the combined linkage data of families RP328, RP349, RP351, RP355 and RP377.

excluding RP377 is that Genehunter does not allow LOD score calculation if there are 2 founders and a non-founder in the pedigree.

Interestingly, in one of the linked families (RP328) a novel crossover was observed with microsatellite marker *D6S1557* (Fig. 3). This may potentially lead to the refinement of the *RP25* genetic interval from the original 16 cM to 2.67 cM between microsatellite markers *D6S257* and *D6S1557*. We believe that the observed crossover in this family and the consequent genetic refinement of the *RP25* interval to 2.67 cM are compatible with our recent identification of ~100-Kb deletion within the refined interval in one of the *RP25* linked families (submitted for publication).

It is interesting to note that there was no common haplotype between the studied families within the *RP25* interval. Ultimately, this can rule out the possibility that the mutation in the causative gene will be due to a common founder effect in all families.

Discussion

In this paper we present validation of linkage to the *RP25* locus in 3 original families as reported by us previously (Ruiz et al. 1998). Furthermore, we show a suggestive evidence of linkage to the same locus in 5 new Spanish arRP families of which one is likely to have refined the *RP25* interval from 16.1 cM to a 2.67 cM between *D6S257* and *D6S1557*.

Previously, positional cloning via family based genomewide linkage analyses has been considered as a powerful approach for the elucidation of both Mendelian and common diseases (Sellick et al. 2004; Farrall & Morris 2005). Likewise, recent advances in SNP identification and genotyping technology have extended the use of dense SNP maps for linkage purposes in place of microsatellite panels (Evans & Cardon 2004).

Recently, targeted linkage analysis and genomewide screening approaches have been combined as useful tools

© 2008 The Authors
Journal compilation © 2008 University College London

for mapping rare and genetically heterogeneous disorders such as RP (Abd El-Aziz et al. 2007; Pomares et al. 2007).

In this study, we undertook a genomewide linkage analysis in three of the originally linked families using the 10K GeneChip array in order to prove *RP25* as the true locus for linkage and to eliminate the possibility of other loci existing across the human genome. The data generated from this analysis revealed the *RP25* locus as the most likely region of linkage and hence supported the original linkage data (Ruiz et al. 1998) and raises an interesting possibility that a disease gene should be present within the *RP25* interval.

Secondly, we have identified 5 out of 18 newly ascertained Spanish arRP families as linked to the *RP25* locus using *RP25* targeted linkage analysis. This has further confirmed the high prevalence of *RP25* which, would increase from 13.7% (Ruiz et al. 1998) to 27.7% among the Spanish population. In addition, a novel crossover in one of the newly linked families, RP328, has further refined the *RP25* interval from 16.1 cM to 2.67 cM. In parallel, we have identified a deletion within the 2.67 cM interval in one of the previously described *RP25*-linked families by using comparative genomic hybridization (CGH) analysis (submitted for publication) therefore highlighting the significance of the current finding.

In summary, we have confirmed the original linkage to the *RP25* locus, and identified five new Spanish families likely to be linked to the same locus. Additionally, a refined interval of 2.67 cM between *D6S257* and *D6S1557* has been established as a probable disease interval for *RP25*. This will help in prioritizing the genes within the refined interval for mutation screening which will eventually lead to cloning of the *RP25* gene. It is interesting to note that the previous and the current genetic data could not identify any regions of common haplotype between affected individuals from different families from Spain across the *RP25* interval. Different mutations within one gene are therefore expected. On the other hand, more than one gene responsible for the RP phenotype in different families may exist. The identification of the gene and/or the genes responsible for the RP25 phenotype would consequently complement our knowledge on the pathophysiology of the retinal degenerations and give rise to possibilities for the development of new therapies.

Acknowledgments

We would like to thank the families who participated in the study. This study is funded by Fondo de Investigación Sanitaria, Spain (PI050857), the British Retinitis Pigmentosa Society (Grant ref.GR556), NIHR Biomedical Research Centre for Ophthalmology (BMRC) and Special Trustees of Moorfields Eye Hospital.

References

Abd El-Aziz, M. M., El-Ashry, M. F., Barragan, I., Marcos, I., Borrego, S., Antinolo, G. & Bhattacharya, S. S. (2005) Molecular genetic analysis of two functional candidate genes in the autosomal recessive retinitis pigmentosa, RP25, locus. *Curr Eye Res* 30, 1081–1087.

Abd El-Aziz, M. M., Patel, R. J., El-Ashry, M. F., Barragan, I., Marcos, I., Borrego, S., Antinolo, G. & Bhattacharya, S. S. (2006) Exclusion of four candidate genes, KHDRBS2, PTP4A1, KIAA1411 and OGFRL1, as causative of autosomal recessive retinitis pigmentosa. *Ophthalmic Res* 38, 19–23.

Abd El-Aziz, M. M., El-Ashry, M. F., Chan, W. M., Chong, K. L., Barragan, I., Antinolo, G., Pang, C. P. & Bhattacharya, S. S. (2007) A Novel Genetic Study of Chinese Families with Autosomal Recessive Retinitis Pigmentosa. *Ann Hum Genet* 71, 281–294.

Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. (2001a) GRR: graphical representation of relationship errors. *Bioinformatics* 17, 742–743.

Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. (2001b) Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30, 97–101.

Ayuso, C., Garcia-Sandoval, B., Najera, C., Valverde, D., Carballo, M. & Antinolo, G. (1995) Retinitis pigmentosa in Spain. The Spanish multicentric and multidisciplinary group for research into retinitis Pigmentosa. *Clin Genet* 48, 120–122.

Barragan, I., Marcos, I., Borrego, S. & Antiñolo, G. (2005a) Molecular analysis of RIMI in autosomal recessive retinitis pigmentosa. *Ophthalmic Res* 37, 89–93.

Barragan, I., Marcos, I., Borrego, S. & Antinolo, G. (2005b) Mutation screening of three candidate genes, ELOVL5, SMAP1 and GLULD1 in autosomal recessive retinitis pigmentosa. *Int J Mol Med* 16, 1163–1167.

Barragan, I., Borrego, S., Abd El-Aziz, M. M., El-Ashry, M. F., Abu-Safieh, L., Bhattacharya, S. S. & Antinolo, G. (2007) Genetic Analysis of FAM46A in Spanish Families with Autosomal Recessive Retinitis Pigmentosa: Characterisation of novel VNTRs. *Ann Hum Genet* 71, 281–294.

den Hollander, A. I., Koenekoop, R. K., Mohamed, M. D., Arts, H. H., Boldt, K., Towns, K. V., Sedmak, T., Beer, M., Nagel-Wolfrum, K., McKibbin, M., Dharmaraj, S., Lopez, I., Ivings, L., Williams, G. A., Springell, K., Woods, C. G., Jafri, H., Rashid, Y., Strom, T. M., van der Zwaag, B., Gosens, I., Kersten, F. F., van Wijk, E., Veltmanb, J. A., Zonneveld, M. N., van Beersum, S. E., Maumenee, I. H., Wolfrum, U., Cheetham, M. E., Ueffing, M., Cremers, F. P., Inglehearn, C. F. & Roepman, R. (2007) Mutations in LCA5, encoding the ciliary protein lebercilin, cause Leber congenital amaurosis. *Nat Genet* 39, 889–895.

Dharmaraj, S., Li, Y., Robitaille, J. M., Zhu, D., Mitchell, T. N., Maltby, L. P., Baffoe-Bonnie, A. B. & Maumenee, I. H. (2000) A novel locus for Leber congenital amaurosis maps to chromosome 6q. *Am J Hum Genet* 66, 319–326.

Dracopoli, N. H., Haines, J. L., Korf, B. R., Moir, D. T., Morton, C. C., Seidman, C. E., Seidman, J. G. & Smith, D. R. (1994) *Current protocols in human genetics*, Vol. 1. New York: John Wiley and Sons.

Evans, D. M. & Cardon, L. R. (2004) Guidelines for genotyping in genomewide linkage studies: single-nucleotide-polymorphism maps versus microsatellite maps. *Am J Hum Genet* 75, 687–692.

Farrall, M. & Morris, A. P. (2005) Gearing up for genome-wide gene-association studies. *Hum Mol Genet* 14(Spec No. 2), R157–R162.

© 2008 The Authors
Journal compilation © 2008 University College London

Inglehearn, C. F. (1998) Molecular genetics of human retinal dystrophies. *Eye* 12, 571–579.
John, S., Shephard, N., Liu, G., Zeggini, E., Cao, M., Chen, W., Vasavda, N., Mills, T., Barton, A., Hinks, A., Eyre, S., Jones, K. W., Ollier, W., Silman, A., Gibson, N., Worthington, J. & Kennedy, G. C. (2004) Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites. *Am J Hum Genet* 75, 54–64.
Jimenez-Sierra, J. M., Ogden, T. E. & van Boemel, G. B. (1989) *Inherited Retinal Diseases: A Diagnostic Guide*. St Louis: Mosby.
Khaliq, S., Hameed, A., Ismail, M., Mehdi, S. Q., Bessant, D. A., Payne, A. M. & Bhattacharya, S. S. (1999) Refinement of the locus for autosomal recessive Retinitis pigmentosa (RP25) linked to chromosome 6q in a family of Pakistani origin. *Am J Hum Genet* 65, 571–574.
Lander, E. S. & Botstein, D. (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236, 1567–1570.
Li, Y., Marcos, I., Borrego, S., Yu, Z., Zhang, K. & Antinolo, G. (2001) Evaluation of the ELOVL4 gene in families with retinitis pigmentosa linked to the RP25 locus. *J Med Genet* 38, 478–480.
Marcos, I., Borrego, S. & Antinolo, G. (2003) Molecular cloning and characterization of human RAB23, a member of the group of Rab GTPases. *Int J Mol Med* 12, 983–987.
Marcos, I., Galan, J. J., Borrego, S. & Antinolo, G. (2002) Cloning, characterization, and chromosome mapping of the human GlcAT-S gene. *J Hum Genet* 47, 677–680.
Marcos, I., Ruiz, A., Blaschak, C. J., Borrego, S., Cutting, G. R. & Antiñolo, G. (2000) Mutation analysis of GABRR1 and GABRR2 in autosomal recessive retinitis pigmentosa (RP25). *J Med Genet* 37, E5.
Middleton, F. A., Pato, M. T., Gentile, K. L., Morley, C. P., Zhao, X., Eisener, A. F., Brown, A., Petryshen, T. L., Kirby, A. N., Medeiros, H., Carvalho, C., Macedo, A., Dourado, A., Coelho, I., Valente, J., Soares, M. J., Ferreira, C. P., Lei, M., Azevedo, M. H., Kennedy, J. L., Daly, M. J., Sklar, P. & Pato, C. N. (2004) Genomewide linkage analysis of bipolar disorder by use of a high-density single-nucleotide-polymorphism (SNP) genotyping assay: a comparison with microsatellite marker assays and finding of significant linkage to chromosome 6q22. *Am J Hum Genet* 74, 886–897.
O'Connell, J. R. & Weeks, D. E. (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63, 259–266.
Pomares, E., Marfany, G. Brion, M. J., Carracedo, A. & Gonzalez-Duarte, R. (2007) Novel high-throughput SNP genotyping cosegregation analysis for genetic diagnosis of autosomal recessive retinitis pigmentosa and Leber congenital amaurosis. *Hum Mutat* 22, 1–6.
Ruiz, A., Borrego, S., Marcos, I. & Antiñolo, G. (1998) A major locus for autosomal recessive retinitis pigmentosa on 6q, determined by homozygosity mapping of chromosomal regions that contain γ-aminobutyric acid-receptor clusters. *Am J Hum Genet* 62, 1452–1459.
Ruschendorf, F. & Nurnberg, P. (2005) ALOHOMORA: a tool for linkage analysis using 10K SNP array data. *Bioinformatics* 21, 2123–2125.
Sellick, G. S., Longman, C., Tolmie, J., Newbury-Ecob, R., Geenhalgh, L., Hughes, S., Whiteford, M., Garrett, C. & Houlston, R. S. (2004) Genomewide linkage searches for Mendelian disease loci can be efficiently conducted using high-density SNP genotyping arrays. *Nucleic Acids Res* 32, e164.
Stambolian, D., Ibay, G., Reider, L., Dana, D., Moy, C., Schlifka, M., Holmes, T., Ciner, E. & Bailey-Wilson, J. E. (2004) Genomewide linkage scan for myopia susceptibility loci among Ashkenazi Jewish families shows evidence of linkage on chromosome 22q12. *Am J Hum Genet* 75, 448–459.
Tuson, M., Marfany, G. & Gonzalez-Duarte, R. (2004) Mutation of CERKL, a novel human ceramide kinase gene, causes autosomal recessive retinitis pigmentosa (RP26). *Am J Hum Genet* 74, 128–138.
Woods, C. G., Valente, E. M., Bond, J. & Roberts, E. (2004) A new method for autozygosity mapping using single nucleotide polymorphisms (SNPs) and EXCLUDEAR. *J Med Genet* 41, e101.
Woods, C. G., Cox, J., Springell, K., Hampshire, D. J., Mohamed, M. D., McKibbin, M., Stern, R., Raymond, F. L., Sandford, R., Malik Sharif, S., Karbani, G., Ahmed, M., Bond, J., Clayton, D. & Inglehearn, C. F. (2006) Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *Am J Hum Genet* 78, 889–896.
Zhang, Q., Zulfiqar, F., Xiao, X., Riazuddin, S. A., Ayyagari, R., Sabar, F., Caruso, R., Sieving, P. A., Riazuddin, S. & Hejtmancik, J. F. (2005) Severe autosomal recessive retinitis pigmentosa maps to chromosome 1p13.3-p21.2 between D1S2896 and D1S457 but outside ABCA4. *Hum Genet* 119, 617–623.
Zhao, C., Lu, S., Zhou, X., Zhang, X., Zhao, K. & Larsson. (2005) A novel locus (RP33) for autosomal dominant retinitis pigmentosa mapping to chromosomal region 2cen-q12.1. *Hum Genet* 118, 356–365.

Received: 14 November 2007
Accepted: 2 April 2008

© 2008 The Authors
Journal compilation © 2008 University College London